



〔澳〕陶哲轩(Terence Tao) 著
〔加〕于青林 译 潘承彪 校

解题·成长·快乐
陶哲轩教你学数学



北京大学出版社
PEKING UNIVERSITY PRESS

著作权合同登记图字：01-2007-0729

图书在版编目(CIP)数据

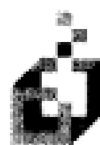
解题·成长·快乐——陶哲轩教你学数学/〔澳〕陶哲轩(Terence Tao)著；〔加〕于青林译. —北京：北京大学出版社, 2009. 7

ISBN 978-7-301-15447-2

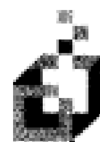
Ⅰ. 解… Ⅱ. ①陶… ②于… Ⅲ. 数学-学习方法
Ⅳ. O1-4: G624.3

中国版本图书馆 CIP 数据核字（2009）第 111614 号

书　　名：解题·成长·快乐——陶哲轩教你学数学
著作责任者：〔澳〕陶哲轩(Terence Tao)　著　〔加〕于青林　译
责 任 编 辑：曾琬婷　孙　琰
标 准 书 号：ISBN 978-7-301-15447-2/G · 2645
出 版 发 行：北京大学出版社
地　　址：北京市海淀区成府路 205 号　100871
网　　址：http://www.pup.cn　电子邮箱：zpup@pup.pku.edu.cn
电　　话：邮购部 62752015　发行部 62750672　编辑部 62752021
出版部 62754962
印 刷 者：北京宏伟双华印刷有限公司
经 销 者：新华书店
650 毫米×980 毫米　16 开本　9.25 印张　100 千字
2009 年 7 月第 1 版　2009 年 7 月第 1 次印刷
定　　价：20.00 元



目　录

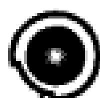


中译版序言

这是一本适合中学生的数学读物。北京大学出版社和潘承彪教授嘱我为它写序言，因为我曾见过少年陶哲轩。1986年第27届国际数学奥林匹克(IMO)在波兰华沙举行。在颁奖大会的休息大厅，我见到了陶哲轩，当时他10岁，是年龄最小的参赛者。他获得了铜牌，又是华裔，我就主动和他交谈。可惜他一句中文也不会说，我就告诉他，他的姓“陶”怎样写。1987年28届IMO在古巴举行，他获得银牌。在颁奖大会上我又见到了他。因相距甚远，我们双方就招手致意。1988年，第29届IMO在澳大利亚举行，他获得了金牌，可惜我未能参加。他在IMO历程上逐年进步，是很受业内人士称道的“神童”式人物。1995年我应邀为澳大利亚数学竞赛出题，我向澳大利亚的朋友问起他的情况，知道他在美国某大学攻读研究生。现在他已经是非常出色的职业数学家。

这本书中的题目，我相信多数他都自己动手做过。在得到了答案之后，对解题过程做了尽情细致的回顾，才能对解答过程分析得十分透彻，从中对解答的思路、技巧有很好的积累。我认为这本书就是这样产生的，而且书中用的中学生语言，值得我们的中学生借鉴。我们的同学解题只要结果，而忽视过程中的精彩心得，因此做出的解答是对的，但不甚漂亮。我认为这也是我们在IMO中金牌已逾百枚，却没有一个人得特别奖(IMO 特别奖是奖给对某一题解法巧妙独特的人)的原因之一吧！

书中多处讲到数学的美。他在第一版序言中说到“……



把一个漂亮的、简洁的几何题用解析几何教科书的方法变成丑陋怪物般的方程来解，就不会给予我们成就感”。其实在澳大利亚中学数学教学中几何内容不多，一个中学生有如此见解是难能可贵的。在学习过程中不断追求数学的美感，才会不时产生思想火花，从而有了精彩独到的思路。

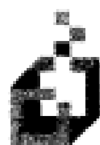
我们的前辈、大数学家华罗庚曾说过“天才在于积累”。陶哲轩的才华亦是不断地积累的，在中学时代就写出这样一本书，就是阶段性积累过程。我也多次听到过国内有这个那个少年有数学才思的学生，一入初中，就把高中数学都学完了，但很遗憾的是后来就没有下文了。这里奉劝一些家长、老师，要尊重教育规律，不要做拔苗助长的事，天才不是天生的。

本书所选择的问题是恰当的，不太难又很有想法，很多题目都选自“环球城市数学竞赛”。这是俄罗斯举办的一个国际数学竞赛，应该说这一竞赛的题目是很好的。现在国内有中译本《环球城市数学竞赛问题与解答(I, II)》(北京：开明出版社，2004)，有兴趣者亦可一试之。



裘宗沪

2009 年 7 月初



原版第二版序言

这本书写于15年前，对于今天的我来说等于半个人生以前。在成长的日子里，我离家远赴异国他乡，考取研究生，教书，撰写研究论文，辅导研究生，结婚，并有了一个儿子。显然，现在我对生活与数学的理解，较之于15岁时改变了很多。我已有很长时间没有涉足解题竞赛了，因此如果我今天来写这样题材的书，那么将会和你正在读到的很不一样。

数学是一门涉及多方面的学问，我们关于它的经验和鉴赏力会随着时间的推移与经历的丰富而变化。当我是小学生时，形式运算的抽象美及其令人惊叹的、通过简单法则的重复而得出非凡结果的能力吸引了我；当我是高中生时，通过竞赛，我把数学当做一项运动，并享受解答设计巧妙的数学趣味题(正像本书中的问题一样)和揭开每一个奥秘的“窍门”时的快乐；当我是大学生时，初次接触到构成现代数学核心的丰富、深刻、迷人的理论和体系，使我顿起敬畏之心；当我是研究生时，我为拥有自己的研究课题而感到骄傲，并从对以前未解决的问题提供原创性证明的过程中得到无与伦比的满足。直到自己开始作为一名研究型数学家的职业生涯后，我才开始理解隐藏在现代数学理论和问题背后的直觉力及原动力。当意识到无论多么复杂和深奥的结果往往都是由非常简单，甚至是常识性的原理导出时，我感到欣喜。当抓住这些原理中的一个，且突然领会到它是如何照亮一个巨大的数学体系并赋予其活力时，“啊！”脱口而出，这真是令人惊奇的非凡体验。然而，仍有很多方面的数学有待发现。直到最

近，当我了解了足够多的数学领域后，才开始理解整个现代数学的努力方向及其与科学和其他学科的联系。

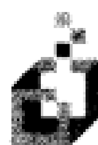
由于本书是我开始职业数学生涯之前完成的，当时我并不具备现在的洞察力和经验，因此书中许多地方的写法具有某种无知，甚至是幼稚的东西。我并不想太多地改变它们，因为年轻时的我比现在的我更能融入高中生的解题世界。然而，我对本书做了若干结构上的调整：用 LaTex 编排格式；把材料组织得我个人认为更有逻辑性；修改那些用词不准确、不当、混淆或结构松散的部分。我还增加了习题的数量。某些地方，内容有些过时(例如费马(Fermat)大定理现在已有了严格的证明)。现在我也意识到，书中的有些问题可以用更便捷、更简洁的“先进”数学工具来解决。但本书的目的并不是对问题提供最简洁的答案或给出最新的结论综述；而是要指明，刚接触一个数学问题时，我们应该如何去处理它，如何通过努力从不同角度尝试一些想法和排除另一些想法，以及如何通过有计划地处理，最终得到一个满意的解答。

我非常感谢 Tony Gardiner 对本书再版所给予我的鼓励和支持，以及我父母多年来的全力支持。我也被所有的朋友和这些年来我遇到的读过本书第一版的人所深深感动。最后(但并非不重要的)，我要特别感谢我的父母和 Flinder 医疗中心的计算机技术人员的支持，是他们从我老旧的苹果计算机(Macintosh Plus) 中复原了本书 15 年前的备份电子版！

陶哲轩

美国加州大学洛杉矶分校(UCLA)数学系

2005 年 12 月



原版第一版序言

古希腊哲学家普罗克洛斯(Proclus)曾说过："这，就是数学：她提醒你灵魂有不可见的形态；她赋予自己的发现以生命；她唤醒悟性，澄清思维；她照亮了我们内心的思想；她涤尽我们有生以来的蒙昧与无知……"[0]

而我喜欢数学，因为她有趣。

数学问题或智力题，对于现实中的数学(即解决实际生活问题的数学)是十分重要的，就如同寓言、童话和奇闻逸事对年轻人理解现实生活的重要性一样。已被其他人发现了优美解法的数学问题是"净化了的"数学，因为问题的表面东西已被剥去，展现出了它的有趣且(有望是)发人深思的形式。如果把学习数学比做勘探金矿，那么解决一个好的数学问题就近似于为寻找金矿而上的一堂"捉迷藏"课：你要去寻找一块金子，你知道这块金子是什么样的，它就在附近的某个地方，要到达那个地方不是太困难，是在你的能力所及的范围内，同时适当地给了你去挖掘它的合适工具(例如已知条件)。因为金子隐藏在一个不易发现的地方，要找到它，比随意挖掘更重要的是正确的思路和技巧。

在本书中，我将解决若干具有不同难度及从不同数学分支中选择的问题。标星号 * 的问题是较难的，因为，它们要么需要某些较深的数学知识，要么需要某些更巧妙的想法；标双星号 ** 的问题难度更大。有些问题后会附带一些习题，它

⓪, ①, ②, …表示校者所加的注，这些注全在本书最后的校后记及校注中。

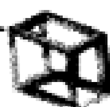
们能用类似的方法解决或涉及类似的数学知识。在解这些题的同时，我将试图阐明解题的一般技巧。解题的两个要素——经验和知识，是不容易被写进书里的：要想获得它们，必须经历时间的磨炼。但是，书里有许多不需要多少时间就可学会的较简单的技巧；有一些分析问题的方法，它们有助于我们较易找到合理可行的处理问题的方案；也有一些系统的分类方法，利用它们可以把一个问题简化为若干相关联的较简单的子问题。然而，解答问题并不是事情的全部。让我们再来看寻找金子的例子，与仔细测量，并用一点儿地质学知识进行小规模的挖掘相比，如露天采矿似地用推土机把邻近地块统统挖一遍就更显得十分笨拙了。一个解法应该相对简洁，容易理解，并且有望达到优美。同时，它也应该有发现的乐趣。把一个漂亮的、简洁的几何题用解析几何教科书中的方法变成丑陋怪物般的方程来解，就不会有只用两行向量的解法所给予我们的成就感。

作为一个典型例子，请看以下欧几里得几何中的一个标准结果：

一个三角形的三条边的垂直平分线是共点的。

这一简洁的“一句话”命题可以用解析几何方法来证明。请读者试着自己花几分钟(几小时？)做一做，然后再看下面的解法：

证明 令这个三角形为$\triangle ABC$。设点P是边AB和AC的垂直平分线的交点(图 1)。因为点P在AB的平分线上，所以$|AP|=|PB|$；同样地，因为点P在AC的平分线上，故$|AP|=|PC|$。结合这两个式子，得到$|BP|=|PC|$。这就意味着点P必须在BC的平分线上。所以，这三条垂直平分线共点(顺便指出，点P是$\triangle ABC$的外接圆的圆心)。 □



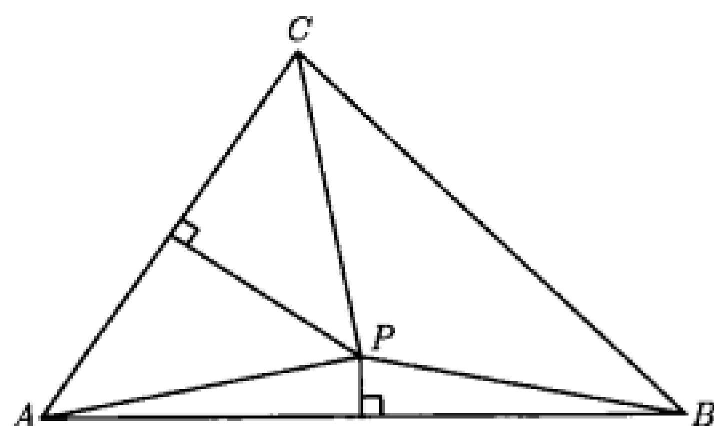


图 1

下面的简图(图 2)说明了如果点 P 在 AB 的垂直平分线上，为什么有结论$|AP|=|PB|$：用两个全等三角形就把这说清楚了。

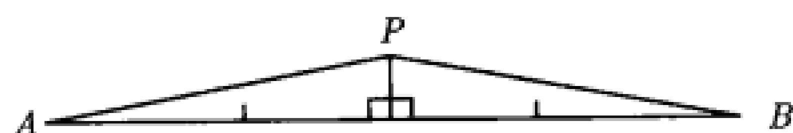


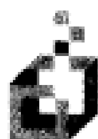
图 2

这种解法——把一些显然的事实相互结合在一起，导出一个不太显而易见的结论——正是数学之美的一部分。我希望你们也将欣赏到这种美！

致谢

感谢 Peter O'Halloran, Vern Treilibs 和 Lenny Ng 所提供的题目和建议。特别感谢 Basil Rennie 所做出的修正和有独创性的便捷解法。最后也感谢我的家庭所给予我的支持、鼓励，纠正我的拼写错误以及鞭策我完成写作计划。

书中几乎所有的题目都出自已出版的数学竞赛习题集；正文中标明了它们的出处，完整的信息见本书的参考文献。我也采用了少量从朋友处或其他数学出版物中获得的题目，这些则没有标明出处。



第一章　解题策略

千里之行，始于足下。

——老子

无论你是否认同这句格言，解答一个问题都起始，并继续，最终结束于简单、合乎逻辑的步骤。但是，只要有敏锐的目光，并以稳健的步伐朝着明确的方向前进，那么千里之行就将远远不需要走千千万万步。数学作为抽象的学问，并没有外部约束，人们总可以从头开始，尝试新的对策或随时返回前一步。而别的学问不一定有这样的灵活性(例如，当你回家迷路时要寻找回家的路)。当然，这并不能使解题变得容易；否则，这本书会薄很多。然而这样的特点却使解题变得可能。

有一些正确解题的一般策略和想法，波利亚的经典文献(Polya, 1957)谈到了其中的很多种。我们会在下面讨论某些策略，同时会简要说明其中每种策略如何运用于下面的问题：

> **问题 1.1**　一个三角形的三条边长构成公差为 d 的等差数列，三角形的面积为 t。求这个三角形的边长和角度。

理解问题。这个问题属于哪种类型呢？数学中的问题主要分三类：

- “证明……”或“推算……”型问题。这类问题要求证明某个命题成立或推算某个表达式的值。
- “求……(值)”或“求所有的……(值)”型问题。这类

问题要求找出满足某些条件的一个或所有的值。

• "是否存在……"型问题。这类问题要求证明一个命题或给出一个反例。

问题的类型决定了解题的基本方法或方式，所以它至关重要。在"证明……"或"推算……"型问题中，从给定的信息入手，其目的是根据事先给出的信息推导出某个命题或计算出某个表达式的值。由于这类问题有清晰的目标，所以通常比另外两类问题来得容易。"求……(值)"型问题更依赖运气，通常要先猜一个相近的答案，再做些小的调整，使它更接近于正确答案；或者先修改题目的要求，使之更容易满足，再考虑原来的要求。"是否存在……"型问题通常是最难的，因为我们必须先判断讨论的对象是否存在，再提供证明或举出反例。

当然，并不是所有的问题都可以这样简单地归类。但通常问题的类型将提供解题的基本策略。例如，要解决这样的问题"在这座城市里找一个今晚可以睡觉的旅馆"，就应先把要求改成如"找一个在 5 km 以内的、有空闲房间的旅馆，且一晚房费不超过 100 美元"，然后采用排除法来找。这种策略比证明这样的旅馆存在或不存在要好，也可能比先随便选一家旅馆，然后证明是否适合休息要好。

在问题 1.1 这个"推算……"型问题中，需要在给定若干变量的情况下求出几个未知量。这就提示我们用代数方法建立多个联系 d, t 以及三角形的三条边和三个角的方程，并最终求解未知量，而不是用几何方法。

理解题目所给出的信息。问题中给出了什么信息呢？通常，一个问题会提到若干个满足某些特定条件的对象。为了理解这些信息，需要观察它们和给定的要求之间是如何相互作用的。把注意力集中在选择恰当的技巧和符号上，这对解



决问题很重要。在我们的例题中，信息包括一个三角形、它的面积以及它的三条边长构成公差为 d 的等差数列。因为有了一个三角形并考虑其边长和面积，所以我们就需要用有关三角形的边、角和面积的定理来解决问题，例如用正弦定理、余弦定理和面积公式等。我们还需要用符号来表示等差数列，例如三角形的边长可表示为 a, $a+d$ 和 $a+2d$。

理解题目所要求的目标。题目要求的目标是什么？也许是求一个值，证明一个命题，或决定一个具有某种特性的对象的存在性，等等。如同在“理解问题所给出的信息”部分所提到的那样，了解目标有助于我们集中精力选择最合适的解题工具，也有助于我们建立一个战术性的目标，使我们更接近于问题的解。在这个问题中，目标是“求这个三角形所有的边长和角度”。如前所述，这意味着我们需要有关三角形边长、角度的定理和公式。而我们的战术性目标是“找到有关三角形边长和角度的关系式”。

选择恰当的符号。我们理解了题意和目标后，还需要把它们用尽可能简单的形式有效地表达出来。这通常涉及前面所讨论的两种策略。在我们的例题中，已经考虑到建立有关 d, t 以及三角形的边长和角度的方程。我们需要用变量表示三角形的边长和角度：可以设边长为 a, b, c；而角度记做 α, β, γ。然而我们可以用题目所给出的信息来简化这些符号：由于知道三条边长呈等差数列，因此可以用 a, $a+d$ 和 $a+2d$ 取代 a, b 和 c。如果令边长为 $b-d$, b 和 $b+d$，使之对称，这样会更好些。这种符号表示唯一的缺陷是，b 必须大于 d。但进一步想，我们发现这不是一个真正的限制。实际上，$b>d$ 给我们提供了额外的信息。我们也可以做较大调整，把角度记为 α, β, $180°-\alpha-\beta$。但是这种表示不好看，也不对称，所以最

好保持原来的符号，不过要记住 $\alpha+\beta+\gamma=180°$。

用选定的符号表达你所知道的信息，并画一个示意图。 把所有的信息写在纸上，有三点好处：

(a) 解题时，便于参考；

(b) 陷入困境时，可以盯着纸进行思考；

(c) 把知道的写下来，这个过程本身可以激发新的灵感和联想。

但是请注意，不要写下过多的信息和细节。一种折中的办法是着重强调那些你认为最有用的事实，而把那些令人怀疑的、冗杂的或异想天开的想法写在另一张草稿纸上。下面是从这个问题想到的一些方程和不等式：

- (自然约束) $\alpha, \beta, \gamma, t>0$ 和 $b>d$。不失一般性地，我们还可以假设 $d\geqslant 0$。
- (三角形的三个角之和) $\alpha+\beta+\gamma=180°$。
- (正弦定理) $\dfrac{b-d}{\sin\alpha}=\dfrac{b}{\sin\beta}=\dfrac{b+d}{\sin\gamma}$。
- (余弦定理) $b^2=(b-d)^2+(b+d)^2-2(b-d)(b+d)\cos\beta$，等等。
- (三角形面积公式)

$$t=\frac{1}{2}(b-d)b\sin\gamma=\frac{1}{2}(b-d)(b+d)\sin\beta=\frac{1}{2}b(b+d)\sin\alpha$$ 。

- (Heron 公式) $t^2=s(s-b+d)(s-b)(s-b-d)$，这里 $s=\dfrac{1}{2}[(b-d)+b+(b+d)]$ 是半周长。
- (三角不等式) $b+d\leqslant b+(b-d)$。

这些事实中，很多可能被证明是无用的或分散注意力的。但是利用某种判断法则，我们可以把有用的事实与无用的事实分离开来。由于我们的目标和信息都以等式的形式出现，



所以等式看上去比不等式更有用。另外，因为半周长可简化为 $s = 3b/2$，所以 Heron 公式看起来有望用得上。因此，我们把“Heron 公式”作为可能有用的事实重点标出来。

当然我们也可以画一张示意图(图 1)。通常这对于几何问题十分有用，但在我们的例题中这样的图似乎提供不了更多的信息。

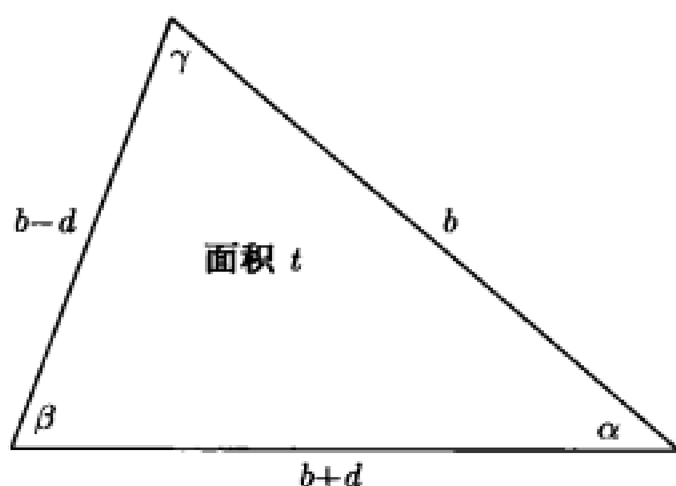


图　1

对问题稍做修改。可用很多种方法来修改问题，使其更容易处理，如：

(a) 考虑该问题的一种特定情形，例如极端情形或退化情形；

(b) 解决该问题的一种简化情形；

(c) 设计一个包含该问题的猜想，并试图先证明它；

(d) 导出该问题的某个推论，并试图先解决它；

(e) 重新表达该问题(例如用反证法证明其逆否命题，或者尝试其某种替代形式)；

(f) 研究类似问题的解；

(g) 推广该问题。

当你对一个问题无从下手时，上述方法是有帮助的，因为解决一个更简单的相关问题有时可以揭示解决原问题的思

路。同样地，考虑极端情形和解决带有附加假设的问题，也可以对问题的一般情形的解法有所启发。但是需要注意，特定情形本质上是特殊的，某些漂亮的技巧可以用来处理它们，却对一般情形毫无用处。这往往发生在特定情形过于特殊时。从适当地修改假设入手，可以保证你始终与原问题的本质尽可能地接近。

在问题 1.1 中，我们可以从 $d=0$ 的特定情形开始。这时，我们需要求出面积为 t 的等边三角形的边长。用标准方法可计算得到 $b=2\sqrt{t}/\sqrt[4]{3}$ 。这说明一般解也应包括平方根或 4 次方根。可是，这个尝试并没有提供解决原问题的思路。别的类似的尝试也收获不大。这意味着，我们需要某种强有力的代数工具来解决这一问题。

对问题做较大修改。在这种大胆的策略中，我们对问题做出重要修改，例如去掉题目给出的条件，交换给出的条件和要求的目标，或者否定目标(例如尝试否定命题，而不是证明原命题)。我们不断尝试，直到找到问题的突破口为止，然后确定哪里是突破的关键，这样就确定了给定条件的关键所在以及难点所在。这种练习也有助于培养判断哪种策略可行而哪种策略不可行的直觉。

在我们这个特定的例题中，可以用四边形、圆等来代替三角形，只是收效甚微：问题变得更复杂了。但是，另一方面，可以看出，这个问题只需要求出三角形的边长和角度，而并不需要知道三角形所在的位置。因此，这进一步使我们确信，应把注意力放在三角形的边长和角度(即 $a, b, c, \alpha, \beta, \gamma$)上，而不是坐标几何或其他类似的方法。

我们可以忽略题目要求的某些目标，例如只求出三角形的三条边长，而不必求出三角形的角度。因为我们知道，根



据正弦和余弦定理，三角形的角度也就可以确定了。所以，只需求出三角形的边长。由于我们知道边长为 $b-d$, b 和 $b+d$，因此只要求出 b 就解决了问题。

我们也可以忽略题目给出的某些信息，例如公差 d，但是这样做，我们会发现可能得到不止一个解，而且没有足够的信息来解决问题。同样地，忽略面积 t 也不能确定一个解。(有时可以忽略**部分**信息，例如只要求面积大于或小于某个阈值 t_0，但这会变得更复杂。因此首先要抓住简单的选择来尝试。)

问题的逆问题(交换条件和目标) 也可能引发一些有趣的想法。假定有一个三角形，其三条边长呈公差为 d 的等差数列，缩放它(或对它进行其他变换)，直到它的面积变为 t。这个过程可以想象为在保持三条边长的公差的同时缩放三角形并使之变形。类似地，也可以考虑具有固定面积 t 的所有三角形，从中构造一个，使其三条边长满足等差数列的要求。这些想法也许最后会成功，但我将给出这个问题的另一种解法。请记住，一个问题可以有不止一种解法，而且并不存在一种解法是绝对最好的。

证明与我们的问题相关的结果。题目给出的条件是要用到的，因此我们应把每一个条件拿来试一试，看看是否能产生更有意义的信息。在试图证明主要结论或求解答案的过程中，证明一些小结果也许对后面有所帮助。无论如何，不要忘记这些小结果，它们可能在后面会起作用。另外，当你陷入困境时，这些小结果也使你有事可做。

在类似于这个有关三角形的“推算……”型问题中，这种策略不一定有效，但不妨一试。我们的战术性目标是求出 b，它依赖于两个参数 d 和 t。换言之，b 是一个函数：$b=b(d,t)$。

(如果这个符号在几何问题中看起来很奇怪，只是因为在几何中有意忽略对象之间的函数依赖关系。例如，Heron 公式给出了以边长 a,b 和 c 表示的面积 A 的显式形式。换言之，A 可以记做函数 $A(a,b,c)$。) 我们可以证明一些关于函数 $b(d,t)$ 的小结果，例如 $b(d,t)=b(-d,t)$ (因为一个公差为 d 的等差数列总有一个公差为$-d$ 的等差数列与其等价)，或者 $b(kd,k^2t)=kb(d,t)$ (这是由把满足 $b=b(d,t)$ 的三角形扩大 k 倍得到的)。我们甚至可以求 b 对 d 或 t 的导数。在这个特定的问题中，这些策略使我们可以做某些归一化处理，例如令 $t=1$ 或 $d=1$，这也为检验最终的答案提供了一种方法。不过，这些技巧在我们的问题中体现出的优越性不太显著，所以这里不予采用。

简化、充分利用题目所给出的信息，实现战术目标。我们已经引进了符号，建立了若干方程，现在应仔细观察如何实现我们制订的战术目标。在简单的问题中，通常存在标准的解题方法 (例如在高中数学教材中，我们进行过充分讨论的代数简化方法)。这往往是解题过程中最长、最难的部分。然而，如果记住有关的定理、题目给出的信息及其用法，尤其是记住题目要求的目标，就可以避免迷失方向。另外，不要盲目地套用某种已知的技巧或方法，而是应提前考虑一下哪里有可能用到这种技巧。这样我们就可以避免在无助于解决问题的方向上耗费大量精力，从而节省大量时间，在最有助于解决问题的方向上努力。

在问题 1.1 中，我们集中考虑了 Heron 公式，从它可以得到我们的战术目标 b。一旦得到了 b，我们注意到正弦定理和余弦定理就可以用来确定α, β, γ。作为取得进展的更进一步的论据，我们又注意到 Heron 公式与 d 和 t 有关——实质上用到了题目给出的所有数据(我们已经把三角形的三条边



长构成等差数列这一事实体现在选定的符号中了)。总之，用 d, t, b 表示的 Heron 公式变为

$$t^2=\frac{3b}{2}\left(\frac{3b}{2}-b+d\right)\left(\frac{3b}{2}-b\right)\left(\frac{3b}{2}-b-d\right)。$$

我们可以将其简化为

$$t^2=\frac{3b^2(b-2d)(b+2d)}{16}=\frac{3b^2(b^2-4d^2)}{16}。$$

现在我们需要求解 b。上式的等号右边是关于 b 的多项式(把 d 和 t 看做常数)，事实上是 b^2 的二次多项式。于是，我们可以很容易地通过求解二次方程得到 b：如果去掉分母，并把所有项移到等号左边，就得到

$$3b^4-12d^2b^2-16t^2=0。$$

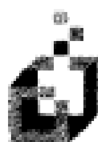
这样，运用二次方程求根公式，有

$$b^2=\frac{12d^2\pm\sqrt{144d^4+192t^2}}{6}=2d^2\pm\sqrt{4d^4+\frac{16}{3}t^2}。$$

因为 b 必须为正的，我们得到

$$b=\sqrt{2d^2+\sqrt{4d^4+\frac{16}{3}t^2}}。$$

作为一种验证，我们可以看到当 $d=0$ 时，上式就等于我们前面的计算结果 $b=2\sqrt{t}/\sqrt[4]{3}$。一旦我们求出三角形的三条边长 $b-d, b, b+d$，就可以由余弦定理推算三个角度α, β, γ。这样，我们就成功了！



第二章 数论中的例子

单数是用来占卜生、死、机缘的。

——莎士比亚，《温莎的风流娘儿们》[1]

也许数论并不是那么神奇，但它还是被神秘的气氛所笼罩。与以等式运算定律为基础的代数不同，数论似乎是从未知源头推导出结论。以拉格朗日(Lagrange)定理(起初是费马的一个猜想)为例，即每个正整数都是四个完全平方数的和(例如 $30 = 4^2+3^2+2^2+1^2$)。从代数的角度看，我们所谈论的是一个非常简单的方程，但因为被限制在整数上，代数法则失灵了。这个结果非常直观，而且数值试验结果表明它似乎是成立的，但就是给不出为什么成立的解释，真是令人气恼！的确，拉格朗日(Lagrange)定理不能用这本书中的初等方法简单证明，而需要用到高斯(Gauss)整数[2]或类似的知识。然而，其他问题却不一定如此深奥。下面是几个简单的例子，它们都与自然数 n 有关：

(a) n 总是和它的 5 次幂 n^5 具有相同的末位数[3]。

(b) n 是 9 的倍数，当且仅当它的各位数字之和是 9 的倍数。

(c) (Wilson 定理) $(n-1)!+1$ 是 n 的倍数，当且仅当 n 是素数。

(d) 如果 k 是一个正奇数，那么 $1^k+2^k+\cdots+n^k$ 可被 $n+1$ 整除。

(e) 存在四个 n 位数的整数(允许补充 0)，每个数的末位



数恰好与其平方的末位数相同。例如，满足这一性质的四个三位数为 000, 001, 625 和 876。

这些命题都可以用初等数论来证明，而且都围绕着**模算术**(modular arithmetic)的基本思想[4]。这会使你体会到代数的威力，但这只适用于有限个整数。顺便提一下，对最后一个命题(e)进行证明的尝试最终可以引出 p 进(p-adic)的概念[5]，它是模算术的一种无穷维形式。

基础数论是封闭的数学乐园。但是基于整数和整除性的应用却广泛、强大得令人惊讶。整除性的概念很自然地引出素数的概念，进而导出因数分解的明确形式，以及在上个世纪末发现的数学的瑰宝之一：素数定理，它可以较精确地给出小于某个给定整数的素数的个数[6]。同时，可以借用整数运算的概念，把模算术从一个整数的子集推广到有限群、环和域的代数。进而，当“数”的概念被推广为无理不尽根、分裂域的元素和复数时，模算术就引出了代数数论。数论是数学的奠基石之一，支撑了数学领域中相当大的一部分。当然，它本身也很有趣。

在开始解题前，让我们复习一些基本的概念：**自然数**就是正整数(我们不把 0 看做是自然数)。全体自然数组成的集合记为 $\mathbf{N}$。**素数**是恰好有两个因数(即这个数本身和 1)的自然数。我们不认为 1 是素数。如果两个自然数 m 和 n 只有公因数 1，则称其为**互素的**。

符号“$x=y \pmod n$”，读做“x 等于 y 模 n”，表示 x 和 y 相差一个 n 的倍数[7]。例如，$15=65 \pmod{10}$。符号“$(\bmod\ n)$”表示我们在**模算术**的意义下工作，这里模数 n 等同于 0。例如，模算术(mod 10) 是在 $10=0$ 条件下的算术，因此

$$65 = 15+10+10+10+10+10$$
$$= 15+0+0+0+0+0 = 15 \pmod{10}。$$

与通常的算术不同是，模算术中不存在不等式，而且其中所有的数均为整数。例如，$7/2 \neq 3.5 \pmod 5$，但因为 $7 = 12 \pmod 5$，故有 $7/2 = 12/2 = 6 \pmod 5$[7]。这种拐弯抹角的除法看起来很奇怪，但事实上我们可以发现这样做并无真正的矛盾，尽管有某些除法是不允许的，正如传统的实数运算中不允许除数为 0 一样。作为模算术中的通用法则，当分母和模数 n 互素时，除法就可以进行[7]。

2.1 位数[8]

前面我们提到过，可以通过考虑自然数的十进位表示中的各位数字之和来了解这个数的性质(比如是否可被 9 整除)。在高等数学中，这种运算并非特别重要(事实证明，直接考察数的本身要比研究它们的十进位表示有效得多)，但在趣味性数学中却十分普遍，有时甚至被赋予某种神秘的含义！当然，位数求和经常出现在数学竞赛题中，正如下面这个例子：

> **问题 2.1** (Taylor，1987，第 7 页) 证明在任意 18 个连续的三位数中至少存在一个整数可以被它的位数和整除[8]。

这是一个有限的问题。因为只存在大约 900 个三位数，所以理论上我们可以动手逐个验证。但是让我们看看是否有省时省力的办法。首先，这个题目的目标看起来有点儿奇特：希望找到可以被其位数和整除的数。我们先用数学公式来表达这个目标，这样处理起来更容易些。为了避免和 abc 混淆，我们可以把一个三位数写成 abc_{10} 的形式，其中 a，b，c 均为阿拉伯数字。需要注意的是，$abc_{10}=100a+10b+c$，而 $abc = a\times b\times c$。沿用标准的符号，用 $a|b$ 表示 a 整除 b，那么，现



在我们要证明的是

$$(a+b+c)|abc_{10}, \tag{1}$$

这里 abc_{10} 是 18 个给定的连续三位数之一。我们是否能够对这个关系式进行变形、简化或有所利用呢？这是可能的，但并不能达到事半功倍的效果(例如得到一个把 a, b 和 c 直接联系起来的有用式子)。事实上，即使用 $100a+10b+c$ 替代 abc_{10}，式(1)依然难以处理。看一看满足式 (1) 的所有解 abc_{10}:

100, 102, 108, 110, 111, 112, 114,

117, 120, 126, …, 990, 999。

这些解看起来是杂乱、随机的。然而，满足条件的解似乎经常出现，使得任意 18 个连续的整数中应该有一个满足式(1)。那么，18 的意义是什么呢？假如 18 不是用来分散注意力的(也许只需要 13 个连续的整数，而 18 只是让你误入歧途)，那么，为什么是 18 呢？也许有人联想到位数和与 9 有很大关系(例如 9 除任何数的余数与 9 除其位数和的余数相同)以及 18 与 9 相关，所以可能存在某种模糊的联系。然而，连续的若干个数和整除性并无直接关系。为了增大解题的机会，我们需要重新表述给定问题或提出一个相关的问题。

我们把注意力放在与数字 9 有关的结果上，就可以发现实际上满足式(1)的数大部分是 9 的倍数，或者至少是 3 的倍数。事实上，在上面列举的数中，只有三个例外(即 100, 110 和 112)，而所有 9 的倍数都满足式(1)。因此，不必直接证明

对于任意 18 个连续的整数，至少有一个满足式 (1);

而是考虑证明

对于任意 18 个连续的整数，存在一个 9 的倍数满足式(1)。

以上想法似乎"打破了"题目给出的信息(18个连续的整数)和目标(满足式(1)的一个整数)之间的"坚冰",因为在18个连续的整数中总是包含一个 9 的倍数(事实上包含两个这样的数),而且从数值验证和数字9所具有的启发性性质中,看起来9的倍数满足式(1)。这种"垫脚石"式的方法是建立两个看似无关的命题之间联系的最佳途径。

这块特定的"垫脚石"(考虑9的倍数)确实有效,但还需要考虑一些例外,以包括所有的情形。事实上,用18的倍数会更好些:

18个连续的整数 → 一个18的倍数 → 式(1)的一个解

做出这种修正基于两方面理由:

- 18个连续的整数总是恰好包含一个18的倍数,但包含两个 9 的倍数。看起来,用 18 的倍数处理这个问题比用 9 的倍数更巧妙、更合适。说到底,如果我们可以用9的倍数来解这问题,那么所问的问题应该是只需要9个连续的整数,而不是18个连续的整数。
- 既然18的倍数只不过是9的倍数的一种特例,对于18的倍数来证明式(1)应该比对于9的倍数更容易。的确,正如我们将看到的,9的倍数并非总是成立(例如909),但18的倍数总是成立的。

总而言之,数值试验表明18的倍数似乎行得通,但这是为什么呢?以216为例,它是18的倍数,它的位数和是9,并且因为18整除216,所以9也可以整除216。再考虑另一个例子,882是18的倍数,而且它的位数和是18,所以882显然可以被 18 整除。花时间研究更多的例子就会看出,18的倍数的位数和总是9或18,所以自然而然地可以整除原来



的整数。综合以上的观察可导出以下的证明。

证明　在 18 个连续的整数中，一定存在一个 18 的倍数，设为 abc_{10}。因为 abc_{10} 也是 9 的倍数，所以 $a+b+c$ 一定是 9 的倍数 (记得 9 的整除法则吗？一个整数被 9 整除，当且仅当它的位数和可以被 9 整除)。因为 $a+b+c$ 的取值范围为 1～27，所以 $a+b+c$ 一定等于 9, 18 或 27。只有当 $abc_{10}=999$ 时，$a+b+c$ 才能等于 27，但它不是 18 的倍数。于是，$a+b+c$ 等于 9 或 18，故有 $a+b+c|18$。由 abc_{10} 的定义知 $18|abc_{10}$ 成立，因此 $a+b+c|abc_{10}$ 得证。 □

请记住，对于涉及位数的问题，直接解法通常行不通。一个复杂的公式应被简化成可以处理的形式。在这个例子中，“任意 18 个连续的整数中一定有一个”可以替换为“任意一个 18 的倍数一定是”，后者结论更弱，但更简单而且与原问题更相关(即与整除性关联)。事实证明，这是一种很好的尝试。还需要记住的是，在有限类型的问题中所用的策略和高等数学中用到的不同。例如，我们不把公式

$$a+b+c|abc_{10}$$

当做典型的数学运算(例如运用模算术)，而是基于所有整数均为三位数的事实来设置 $a+b+c$ 的取值范围(9, 18 或 27)，从而上式变得简单得多：

$$9|abc_{10},\quad 18|abc_{10}\quad 或\quad 27|abc_{10}。$$

的确，尽管逻辑上我们似乎应先把 abc_{10} 表示成代数形式 $100a+10b+c$，但实际上根本不必这样做。这种表示形式只会分散我们的注意力，并不能使问题更容易解决。

最后的注释：事实证明，至少要有 18 个连续的整数，才能保证其中存在一个满足式(1)；17 个整数行不通，例如考虑

559～575 的一组数(我曾经用计算机做过这个例子，并不需要多少数学技巧)。当然，为了解答本题我们不必知道这一事实。

习题 2.1 在一个室内游戏中，"魔术师"请一名观众先想一个三位整数 abc_{10}，然后把 acb_{10}，bac_{10}，bca_{10}，cab_{10} 和 cba_{10} 这五个数加起来，并把求和的结果告诉大家。假定五个数的和为 3194，那么原数 abc_{10} 是多少呢？(提示：先找到一个更适于数学讨论的表达这五个数之和的公式，然后用模算术来得到关于 a, b 和 c 的取值范围。)

问题 2.2 (Taylor, 1989, 第 37 页) 是否存在一个 2 的幂，使其位数重新排列后成为另一个 2 的幂？(首位数不能是 0，例如 0032 是不允许的。)

这个问题涉及 2 的幂以及位数的重排，看起来像一个不可解决的混合体。这是因为：

(a) 位数重排有太多种可能性；

(b) 确定 2 的幂的各个位数并非一件容易的事。

这就意味着我们可能要使用一些不一般的方法。

第一步要做的是猜答案。根据它的出处(即这个问题来自数学竞赛)推测，这不是一个试错法类型的问题，因此答案也许应该是"不存在"。(不过，通过某种极其精巧的构造，可以实现一种巧妙的位数重排，但这样的构造并非易事。所以，先选择一些简单的方案来猜测。如果你猜对了，就可以避免一些无谓的摸索，从而节省大量时间；如果你猜错了，就注定要进行持久而艰苦的努力。这并不意味着你应该把有价值但有难度的解题方案抛在脑后，而是说在深入研究问题之前



要做一个切合实际的评估。)

正如在问题 2.1 中一样，给出位数只是一种分散注意力的障眼法。我们只需知道有关位数和的两个事实：首先是整除性的条件；其次是对取值范围的限制。我们并不想去面对引入一个准确的公式所带来的各种复杂性。这次可能也一样，我们需要通过从位数置换®的过程所得到的结论来简化问题。从纯逻辑的角度来看，因为我们不得不证明更多结论，所以问题会进一步复杂化；但从清晰和简洁程度来看，我们却取得了进展。(为什么要让不好用的信息束缚你呢？它们只会分散你的注意力。)

因此，只要我们把 2 的幂和位数置换的主要性质找出来，运气好的话，就可能找到两者相互矛盾的性质。让我们先着手处理较容易的部分，即 2 的幂，它们是：

$$1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048,$$
$$4096, 8192, 16384, 32768, 65536, \ldots。$$

这里看不出有关位数的太多性质。2 的幂的末位数显然是偶数(1 除外)，但是其他位数是随机的。以整数 4096 为例，它的一个位数为奇数，两个位数为偶数，还包括一个 0。为什么它不能被重排成另一个 2 的幂呢？例如，能否重排为 $2^{4256} = 1523\cdots936$？你会回答：当然不行！这是为什么呢？因为它太大了！这样说的话，与数的大小有关系？是的，2^{4256} 有数千位，而 4096 只有 4 位。啊哈——这样重排位数不会改变位数的个数。(写下任何对解决问题有用的事实，即使它们很简单，不要以为“显而易见”的事实总会在你需要的时候涌现在脑海中。这就像埋藏得不深的金子也要通过寻找并不断地挖掘才能找到。)

能否从这一点儿信息继续我们的推广方案呢？现在，推

广后的问题的是：

> 是否存在一个2的幂，使得另一个2的幂与其具有相同的位数个数？

令人遗憾的是，对于这个问题，我们可以很快地得到一个肯定的答案，例如2048和4096。可见，我们推广得太过份了。(请注意，对上述问题回答“是的”并不意味着对原问题回答“是的”。)再回过头来看问题2.1，仅仅知道“一个18的倍数的位数和一定是9的倍数”并不足以解决这个问题，我们还需要另一个事实，即“三位数的位数和最大是27”。简而言之，我们还没有找到足够多的事实来解决问题。然而，因为我们限制了位数重排的可能性，所以取得了部分的成功。再看4096这个例子，它只能被重排成另一个四位数。在2的幂中只有四个四位数：1024, 2048, 4096和8192。这是因为2的幂总是成倍增大，它们不可能总停留在同一个“数量级”上。事实上，我们很快就可以看到，最多只有四个2的幂可以具有相同的位数个数。(在五个连续的2的幂中，第五个数是第一个数的16倍，所以它的位数一定比第一个数的位数多。)这意味着，对于每个2的幂，最多存在三种其他的排列是2的幂。这样，我们就取得了局部胜利：对于每个2的幂，只有不超过三种可能性需要排除，而不必再考虑无限多种可能性。也许进一步的努力可以把剩下的可能排除掉。

前面提到，当我们进行位数置换时，置换后的数与原数具有相同的位数个数，但逆命题完全不成立，因此位数置换这单一性质本身并不能解决问题。这意味着我们的推广走得太远了，使得解决问题的可能更渺茫。我们回过头来重新观察位数置换，会发现还有其他一些性质也将被保留下来。让我们来看一些例子，仍以4096为例，因为对这个数我们已经



有了一些认识。所有可能的位数置换为

4069, 4096, 4609, 4690, 4906, 4960, 6049, 6094, 6409,

6490, 6904, 6940, 9046, 9064, 9406, 9460, 9604, 9640。

这些数有什么共性呢？它们具有相同的位数集合[8]。这个结论正确无误，但“位数集合”在数学上并不是一个很有用的研究对象(因为有关这个概念的定理和工具不是很多)。然而，**位数和**却是一种经典的数学工具。并且，如果两个数具有相同的位数集合，则它们一定具有相同的位数和。因此我们得到了另一信息：位数置换时保持位数和不变。把这一信息和我们在前面得到的信息结合起来，我们就得到了以下的新命题：

> 是否存在一个 2 的幂，使得另一个 2 的幂与其具有同样多个位数**以及**相同的位数和？

如果这个问题的答案是肯定的，则原问题的答案也是肯定的[9]。这个问题比原问题要容易处理，因为“位数的个数”和“位数和”是标准的数论语言。

让我们记住这些新概念，来考察 2 的幂的位数和，并记住我们的新问题是和它有关的，于是得到表 1。

表　1

2 的幂	位数和	2 的幂	位数和	2 的幂	位数和
$2^0=1$	1	$2^8=256$	13	$2^{16}=65536$	25
$2^1=2$	2	$2^9=512$	8	$2^{17}=131072$	14
$2^2=4$	4	$2^{10}=1024$	7	$2^{18}=262144$	19
$2^3=8$	8	$2^{11}=2048$	14	$2^{19}=524288$	29
$2^4=16$	7	$2^{12}=4096$	19	$2^{20}=1048576$	31
$2^5=32$	5	$2^{13}=8192$	20		
$2^6=64$	10	$2^{14}=16384$	22		
$2^7=128$	11	$2^{15}=32768$	26		

从表 1 中，我们注意到：

• 位数和往往十分小。例如，$2^{17}=131072$ 的位数和只有 14。因为较小的数比较大的数更容易相等，所以实际上较小的数并不表示好运。(如果让 10 个人中的每个人随机选择一个两位数，会有相当大的机会(9.5%)出现一对相同的数。但如果让他们从 10 位数中挑选，则只有非常微小的机会(0.0001%)出现一对相同的数。这简直和中彩票一样不可能。)但是较小的数也有助于找出规律，所以也许并不全是坏消息。

• 某些位数和是相同的，例如 16 和 1024。然而位数和看起来是慢慢增大的：你可以想象 2 的 100 次幂的位数和比 2 的 10 次幂的位数和大。但请记住，我们的条件是具有位数个数相同的 2 的幂，所以考虑位数和相同这种思路帮助不大。

由上述观察可以得出结论：位数和具有某种显而易见的宏观结构——随着幂次 n 的增大而慢慢增大。事实上，对大的 n, 2^n 的位数和极有可能近似于 $(4.5\log_{10}2)n\approx1.355n$ (虽然没有证明)；但它却有一个糟糕的微观结构，即位数的波动太大了。前面我们提到，“位数集合”不便于广泛使用，现在看来“位数和”也没有闪现出什么效果。那么，是否存在这个问题的另一种便于处理的简化形式呢？

我们在前面曾提到，“位数和”是数学中的一种“经典工具”。但是唯一能使其成为真正“制胜”的方法是考虑模 9 的位数和。正如求解问题 2.1 时考虑的一样，我们使用结论：任意整数总是与其位数和模 9 相等。例如，因为 10 等于 1(mod 9)，所以我们可得

$$\begin{aligned}3297&=3\times10^3+2\times10^2+9\times10^1+7\times10^0 \pmod 9\\&=3\times1^3+2\times1^2+9\times1^1+7\times1^1 \pmod 9\\&=3+2+9+7 \pmod 9。\end{aligned}$$



根据这样的考虑，修改后的新问题变为：

> 是否存在一个 2 的幂，使得有另一个 2 的幂与其具有相同的位数个数以及对模 9 有相同的位数和？

利用一个数对模 9 总是等于它的位数和的事实，我们可以把上述问题重新叙述为：

> 是否存在一个 2 的幂，使得有另一个 2 的幂与其具有相同的位数个数以及对模 9 有相同的余数？

请注意，“位数重新排列”、“位数集合”和“位数和”这几个惹人烦的概念就这样被躲开了，新的问题看起来有望得到解决。

让我们对以上列举的 2 的幂的位数和(表 1)进行修改，看看能得到些什么信息(表 2)。

表　2

2 的幂	模 9 的余数	2 的幂	模 9 的余数	2 的幂	模 9 的余数
$2^0 = 1$	1	$2^8 = 256$	4	$2^{16} = 65536$	7
$2^1 = 2$	2	$2^9 = 512$	8	$2^{17} = 131072$	5
$2^2 = 4$	4	$2^{10} = 1024$	7	$2^{18} = 262144$	1
$2^3 = 8$	8	$2^{11} = 2048$	5	$2^{19} = 524288$	2
$2^4 = 16$	7	$2^{12} = 4096$	1	$2^{20} = 1048576$	4
$2^5 = 32$	5	$2^{13} = 8192$	2		
$2^6 = 64$	1	$2^{14} = 16384$	4		
$2^7 = 128$	2	$2^{15} = 32768$	8		

我们需要证明的是，不存在两个 2 的幂具有相同的位数个数及相同的模 9 余数。从表 2 中可以看到，若干个 2 的幂

具有相同的模 9 余数，例如 1, 64, 4096 和 262144，但这四个数并不具有同样多个位数。具有相同模 9 余数的 2 的幂的确是彼此相距较远的，不可能具有同样多个位数。事实上，具有相同模 9 余数的 2 的幂看起来排列得非常有规律。我们可以容易看出，2 的幂的模 9 余数依次每 6 步就重复一次。这个猜测用模算术很容易给出证明：

$$2^{n+6}=2^n2^6=2^n\times 64=2^n \pmod 9,$$

这里我们用到了 $64=1 \pmod 9$。上述结果表明，2 的幂的模 9 余数会如循环小数一样无限地循环重复下去：1, 2, 4, 8, 7, 5, 1, 2, 4, 8, 7, 5, 1, 2, 4, 8, 7, 5, …。这一列循环重复的数还表明，两个具有相同的模 9 位数和的 2 的幂一定相差至少 6 步。但是因为这样两个数至少相差 64 倍，所以它们不可能具有同样多个位数。因此，不存在两个 2 的幂具有同样多个位数以及相同的模 9 位数和。这样，我们就已经证明了修改后的问题。于是我们可以反推过去对原问题做出解答，并写出完整的证明。

证明 假定存在两个 2 的幂，可以通过位数置换而由一个得到另一个。这意味着它们具有同样多个位数以及相同的模 9 位数和。但是模 9 位数和是周期为 6 的一列数，而且在任一给定的周期内没有重复，所以这两个 2 的幂至少相差 6 步。因此，它们就不可能具有同样多个位数，与假设矛盾。 □

通过不断地简化，这个问题中不能用的和不好用的信息被更自然、更灵活和更便于使用的概念所替代。这个简化过程可能具有一点儿偶然性，因为总是存在简化过度或简化不当(简化到歧途上)的可能性。但是在上述问题中，几乎任何处理方法都比翻来覆去地进行位数置换好得多，所以简化不可能使情况变得更糟。也许有时调整和简化会使你白费力气，



但如果你真的束手无策，任何方法都值得一试。

2.2 丢番图方程

丢番图(Diophantus)方程是一类所有变量均为整数的代数方程[①](例如 $a^2+b^2=c^2$ 就是典型的一个)。通常的目标是寻找方程所有的解。一般说来，即使要求所有变量均为整数，方程也会存在不止一个解。求解这类方程，可以用代数方法，也可以用整数除法、模算术和整数的因数分解法等数论中的方法。这里给出一个例子：

> **问题 2.3** (Australian Mathematics Competition, 1987, 第 15 页) 对于非零整数 a 和 b(且 $a+b\neq 0$)，求满足方程
>
> $$\frac{1}{a}+\frac{1}{b}=\frac{n}{a+b}$$
>
> 的所有整数 n。

这是一个标准的丢番图方程，所以我们可以先用乘法消去分母，得到

$$\frac{a+b}{ab}=\frac{n}{a+b},$$

从而有

$$(a+b)^2=nab。\qquad (2)$$

接下来该做什么呢？我们可以忽略 n 而说

$$ab\mid(a+b)^2$$

(这里用了我们曾在问题 2.1 中用过的整除符号“|”)，或者集中考虑 nab 是一个平方数的这一事实。这些想法都不错，但对这一问题不一定有效，因为式 (2) 等号两边联系得不够紧密：左边是平方数；而右边是乘积。

解题时要始终牢记的是，应暂时放弃很有趣但效果不佳的方法，而去尝试更有希望的方法。我们可以先尝试代数方法；如果不成功，再使用数论技巧。把式 (2) 展开并合并同类项，可以得到

$$a^2+(2-n)ab+b^2=0。$$

如果你大胆地用二次方程求根公式，可得

$$a=\frac{b}{2}\left[(n-2)\pm\sqrt{(n-2)^2-4}\right]。$$

上式看起来非常复杂，但实际上我们可以对这种复杂性加以有效利用。我们知道 a, b 和 n 是整数，且上式中有一个平方根，只有平方根中的项 $(n-2)^2-4$ 是一个完全平方数才行。而这就意味着一个平方数减 4 是另一个平方数。这是非常严格的限制。因为在最初的几个平方数之后，两个平方数的间隔都比 4 大，所以我们只需检验几个较小的 n 值。容易发现，$(n-2)^2$ 一定是 4，故 n 为 0 或 4。下面我们分别讨论这两种情形，对每种情形给出相应的解，或者证明这样的解不存在。

情形 1： $n=0$。把 $n=0$ 代回式(2)，我们得到$(a+b)^2=0$，故 $a+b=0$。但是这样在原方程中就出现了无效的 0/0，这是无意义的。所以，n 不能为 0。

情形 2： $n=4$。由式 (2) 得到 $(a+b)^2=4ab$，整理后得到 $a^2-2ab+b^2=0$。对该式进行因数分解，得到 $(a-b)^2=0$，故 a 一定等于 b。这并不产生矛盾：把 $a=b$, $n=4$ 代入式 (2)，原方程成立。

因此，我们的答案是 $n=4$。二次方程求根公式通常很笨拙，通过它求解并不是一种好方法。但因为它引入了一个平方根项，而这一项必须是一个完全平方数，所以有时也能派上用场。



当某个变量出现在指数上，丢番图问题可能变得极其难解。最著名的例子是费马大定理，它断言：对于 $n>2$，$a^n+b^n=c^n$ 不存在自然数解。不过有些涉及指数的问题也比较容易处理，如下面的问题：

> **问题 2.4** (Taylor, 1989，第 7 页)　求出 $2^n+7=x^2$ 的所有解，这里 n 和 x 是整数。

这类问题通常需要反复试验才能找到正确的解决途径。对于丢番图方程，最初等的方法就是模算术和因式分解。模算术可以把整个方程转化为一个恰当的模的关系式，其中模有时是常数(例如(mod 7)或 (mod 16))，而有时是变数(例如(mod pq))；因式分解可以把问题转化成(因子)×(因子)= (易处理的表达式)的形式，这里等号右边可以是常数 (这是最理想的状况)、素数、平方数或其他只有有限种可能因子的项。例如，在问题 2.3 中，我们首先考虑了上述两种方法，但后来倾向于代数方法，而代数方法实际上用到了经过伪装的因式分解技巧(还记得我们最终得到$(n-2)^2-4=$(完全平方数)吗？)。

这里最好先试一试初等技巧，也许可以避免过早陷入怪圈中。我们也可以放弃这两种初等方法，而尝试分析下面的近似方程

$$x=\sqrt{2^n+7}\approx 2^{n/2}。$$

这个方程可能涉及例如连分数、Pell 方程、递归关系等若干高深的数论知识。问题虽可以得到解决，但我们希望找到一种漂亮的 (即省事的) 解法。

如果 n 不是偶数，几乎不可能得到某种有用的因式分解。为此我们可以假定 n 是偶数，考虑两个平方数的差(在丢番图

方程中这是至关重要的技巧)：

$$7 = x^2 - 2^n = (x - 2^m)(x + 2^m),$$

这里 $m = n/2$。于是我们就说 $x - 2^m$ 和 $x + 2^m$ 是 7 的因子，它们一定为 $-7, -1$ 或 $1, 7$。进一步分情形讨论，很快可以证明每种情形都不存在解 (如果我们假定 n 是偶数)。利用因式分解法只能得到 n 必须是奇数这一信息，而并不能告诉我们真正的解是什么以及到底有几个解。

接下来考虑模算术方法。其基本策略是利用模来消去一项或若干项。例如，我们可以把它写成模 x 的方程

$$2^n + 7 = 0 \pmod x$$

或模 7 的方程

$$2^n = x^2 \pmod 7。$$

令人遗憾的是，这些方法都行不通。但是在我们放弃之前，还有一个模数可以一试。我们已经尝试消去含有 7 和 x^2 的项，那么能否消去含有 2^n 的项呢？答案是肯定的。例如，通过选择模 2，当 $n>0$ 时，我们可以得到

$$0 + 7 = x^2 \pmod 2；$$

而当 $n = 0$ 时，可得到

$$1 + 7 = x^2 \pmod 2。$$

这种尝试还是比较理想的，n 已经几乎被消去了。但是问题还没有得到解决，因为等号右边的项 x^2 可以是 0 或 1，我们实际上还没有排除任何可能的取值。为了限制 x^2 的值，我们不得不选择不同的模数。根据这条线索——为了限制上式等号右边 x^2 的值，应考虑尝试选取模 4，而不是模 2：

$$2^n + 7 = x^2 \pmod 4。$$

换句话说，我们得到



$$0+3=x^2\ (\mathrm{mod}\,4)\quad (\text{当 } n>1 \text{ 时}), \tag{3}$$

$$2+3=x^2\ (\mathrm{mod}\,4)\quad (\text{当 } n=1 \text{ 时}), \tag{4}$$

$$1+3=x^2\ (\mathrm{mod}\,4)\quad (\text{当 } n=0 \text{ 时})。 \tag{5}$$

因为 x^2 一定是 0(mod 4) 或 1(mod 4)，故式(3)成立的可能性可以排除。这就意味着 n 只能是 0 或 1。简单的验证表明，n 只能是 1，而 x 一定是+3 或−3。

当求解“找出所有解”的丢番图方程时，主要想法是除了有限个可能取值外排除所有其他可能性。这就是(mod 7)和(mod x)行不通的另一个原因，因为它们会排除所有的取值。而运用(mod 4)时，我们排除了绝大部分可能性而只剩下少数几种可能情形。

> **习题 2.2** 找出最大的正整数 n，使得 n^3+100 可以被 $n+10$ 整除。(提示：用(mod $n+10$)，并利用 $n=-10$ (mod $n+10$)的事实消去 n。)

2.3 幂和

> **问题 2.5** (Hajós 等，1963，第 74 页) 证明对任意非负整数 n，整数 $1^n+2^n+3^n+4^n$ 可以被 5 整除当且仅当 n 不能被 4 整除。

乍一看，这个问题有点儿令人望而生畏，因为式子 $1^n+2^n+3^n+4^n$ 可能会让人想起费马大定理，而众人皆知它是不可解的。但是实际上我们的问题容易得多。我们希望证明某个特定的数可以(或不可以)被 5 整除。除非直接运用因式分解有明显优势，否则我们使用模算术方法。(也就是说，证明对于不能被 4 整除的 n，$1^n+2^n+3^n+4^n=0$ (mod 5)；否则

$1^n+2^n+3^n+4^n \neq 0 \pmod 5$。)

因为这里涉及的数较小，所以可以手工计算 $1^n+2^n+3^n+4^n \pmod 5$的某些值。最好的处理方法是先分别求出 $1^n \pmod 5$, $2^n \pmod 5$, $3^n \pmod 5$和 $4^n \pmod 5$，然后再相加，见表 3。

表 3

n	$1^n \pmod 5$	$2^n \pmod 5$	$3^n \pmod 5$	$4^n \pmod 5$	$1^n+2^n+3^n+4^n \pmod 5$
0	1	1	1	1	4
1	1	2	3	4	0
2	1	4	4	1	0
3	1	3	2	4	0
4	1	1	1	1	4
5	1	2	3	4	0
6	1	4	4	1	0
7	1	3	2	4	0
8	1	1	1	1	4

显然在表 3 中存在某种周期性。事实上 $1^n \pmod 5$, $2^n \pmod 5$, $3^n \pmod 5$ 和 $4^n \pmod 5$ 的变化都以 4 为周期。为了证明这一猜测，我们只需简单地运用周期性的定义。例如，考虑 3^n。以 4 为周期就意味着

$$3^{n+4}=3^n \pmod 5。$$

因为 $81=1 \pmod 5$，所以上式很容易得证：

$$3^{n+4}=3^n \times 81=3^n \pmod 5。$$

类似地，我们可以证明 $1^n \pmod 5$, $2^n \pmod 5$ 和 $4^n \pmod 5$ 的周期也是 4。这意味着$(1^n+2^n+3^n+4^n) \pmod 5$的周期也是 4。这样，我们的问题只需对 $n=0, 1, 2, 3$ 的情形进行证明，



因为其他情形均可由周期性推出。但是我们已经证明了命题对于这些情形都是成立的(观察表 3)，所以证明结束。(顺便提一下，如果我们假定 n 是奇数，还有更初等的证明方法，即对某些项进行简单的配对和抵消。)

只要所证的方程包含一个参数(这个问题中的 n)，周期性总是一种便捷的工具。我们不必对参数的所有取值进行验证，而只需验证一个周期(例如 $n = 0, 1, 2$ 和 3)就足够了。

> **习题 2.3**　如果 x 和 y 是整数，证明方程 $x^4 + 131 = 3y^4$ 没有解。

下面我们转过来考虑一个更棘手的有关幂和的问题：

> ****问题 2.6** (Shklarsky 等, 1962，第 14 页)　设 k 和 n 是自然数，k 是奇数。证明：和式 $1^k + 2^k + \cdots + n^k$ 可以被 $1 + 2 + \cdots + n$ 整除。

顺便提一下，这个问题是伯努利(Bernoulli)多项式的典型练习(或余数定理的某些巧妙应用)。伯努利多项式是数学中一个很有趣的部分，有多种应用。但是在没有伯努利多项式 (或黎曼(Riemann) ζ 函数)这样的有效工具时，我们将只能回到朴素、古老的数论上。

首先，我们知道 $1+2+\cdots+n$ 可以被写成 $n(n+1)/2$ 的形式。我们该用哪种形式呢？前一种更具有美感，但在整除性问题中不太好用(如果可以的话，把除数表示为一个乘积，总比表示为一个和式好用)。如果 $1^k + 2^k + \cdots + n^k$ 可以因式分解出包含 $1 + 2 + \cdots + k$ 的项，也许还有用，但这样的分解并不存在 (至少不是显而易见的)。另外，如果通过某种方式可以把 $1+2+\cdots+n$ 的整除性和 $1+2+\cdots+(n+1)$ 的整除性联系起

来，归纳法也许行得通，但这里看起来可能性也不大。因此，我们将尝试 $n(n+1)/2$ 这个表达式。

我们的目标可以用模算术 (这是证明一个数整除另一个数的最灵活的方法) 表示为

$$1^k+2^k+\cdots+n^k=0\ (\mathrm{mod}\ n(n+1)/2)。$$

让我们忽略 $n(n+1)/2$ 中的“2”，试图证明以下形式的命题：

(因子 1)×(因子 2)|(表达式)。

如果两个因子是互素的，则我们的结论等价于分别证明

(因子 1)|(表达式)　和　(因子 2)|(表达式)。

这使证明变得更简单。因为除数变小，证明整除性会变得更容易。但是还有一个讨厌的“2”碍事。为了对付它，我们将根据 n 是偶数和奇数分成两种情形●。这两种情形十分相似，这里我们只讨论 n 是偶数的情形。在这种情形中，我们可以把 n 记为 $n=2m$(这是为了避免被下列方程中讨厌的项“$n/2$”分散注意力——这类顺手的清理工作将有助于解答的顺利进行)。用 $2m$ 来替换所有的 n，得到

$$1^k+2^k+\cdots+(2m)^k=0\ (\mathrm{mod}\ m(2m+1))。$$

m 和 $2m+1$ 是互素的，上式就等价于

$$1^k+2^k+\cdots+(2m)^k=0\ (\mathrm{mod}\ (2m+1))$$

和

$$1^k+2^k+\cdots+(2m)^k=0\ (\mathrm{mod}\ m)。$$

让我们先处理(mod $(2m+1)$)的部分。它与问题 2.5 十分相似，但因为我们知道 k 是奇数，所以要容易些。当模为 $2m+1$

● 另一种方法是等号两边同时乘以 2，这样我们需要证明 $2(1^k+2^k+\cdots+n^k)=0\ (\mathrm{mod}\ n(n+1))$。其后续讨论和书中给出的方法基本等价。



时，$2m$ 等于-1，$2m-1$ 等于-2，等等。所以我们的表达式 $1^k+2^k+\cdots+(2m)^k$变为

$$1^k+2^k+\cdots+(m)^k+(-m)^k+\cdots$$
$$+(-2)^k+(-1)^k \pmod{(2m+1)}。$$

我们这样做是为了可以进行某些抵消。k是奇数，故 $(-1)^k=-1$，所以 $(-a)^k=-a^k$。这样做的结果是，上面和式中的各项可以配对并消去：2^k和 $(-2)^k$抵消，3^k和 $(-3)^k$抵消，等等。最后只余下一项 $0 \pmod{(2m+1)}$，这正是我们需要证这明的。

然后我们需要处理$\pmod{m}$的部分。也就是说，我们需要证明

$$1^k+2^k+3^k+\cdots+(m-1)^k+m^k+(m+1)^k$$
$$+\cdots+(2m-1)^k+(2m)^k=0 \pmod{m}。$$

在模为 m 时，上式中的某些项可以被简化：m 和 $2m$ 都等于 0，$m+1$ 等于 1，$m+2$ 等于 2，等等。所以上式等号左边可以简化为

$$1^k+2^k+3^k+\cdots+(m-1)^k+0^k+1^k$$
$$+\cdots+(m-1)^k+0 \pmod{m},$$

其中若干项出现了两次，通过合并同类项(并舍弃 0)，我们得到

$$2(1^k+2^k+3^k+\cdots+(m-1)^k) \pmod{m}。$$

于是我们几乎可以做与 $\pmod{(2m+1)}$ 情形同样的处理，只是当 m 是偶数时有点儿麻烦。如果 m 是奇数，上式可以重新表达为

$$2(1^k+2^k+3^k+\cdots+((m-1)/2)^k+(-(m-1)/2)^k$$
$$+\cdots+(-2)^k+(-1)^k) \pmod{m}。$$

然后可以如前进行同样的抵消处理。但是，如果 m 是偶数(设

$m=2p$)，则存在一个中间项 p^k，它不能和任何项相互抵消。换句话说，在这种情形中，表达式并不能立即变成 0，但经过抵消后变成

$$2p^k \pmod{2p}。$$

上式显然等于 0。所以，无论 m 是奇数还是偶数，我们都已经证明了：若 n 是偶数，则 $1^k+2^k+\cdots+n^k$ 可以被 $n(n+1)/2$ 整除。

习题 2.4 当 n 是奇数时，给出上述问题的完整证明。

下面让我们转向一种特殊类型的幂和问题——倒数和。

问题 2.7 (Shklarsky 等, 1962，第 17 页) 设 p 为大于 3 的素数。证明分数和

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}$$

的既约分子可以被 p^2 整除。例如，当 $p=5$ 时，分数

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=\frac{25}{12}$$

的分子显然可被 5^2 整除。

这是一个“证明……”型问题，而不是“求……(值)”或“是否存在……”型问题，因此这样的结论不是完全不可能的。然而，我们需要证明一个关于分数和通过约分后得到的既约分数的分子的结论，这不是一件容易的事！我们需要把这个分子转化成某种更标准的形式，例如一个代数表达式，以便更容易处理。而且，这个问题不只要求被素数整除，还要求被素数的平方数整除。这大大增加了问题的难度。为了使这个问题容易得到解决，我们将以某种方法把问题转化成只与



素数整除性有关的问题。

通过对问题的观察，我们确定了以下目标：

(a) 把分子表达成一个便于处理的数学表达式；

(b) 把 p^2 的整除性问题转化为更简单的问题，例如变成 p 的整除性问题。

让我们先处理目标(a)。首先，得到分子很容易，但得到的分子可能不是一个既约分子。找到公分母通分，把分子加起来，我们得到

$$\frac{[2\times 3\times\cdots\times(p-1)+1\times 3\times\cdots\times(p-1)+\cdots+1\times 2\times 3\times\cdots\times(p-2)]}{(p-1)!}。$$

假定我们可以证明上式中的分子被 p^2 整除，那么怎样才能帮助我们证明约分后得到的既约分子也可以被 p^2 整除呢？既约分子又是什么呢？它是通过对原分子和分母进行约分而得到的。约分是否会破坏 p^2 的整除性呢？如果 p 的某个倍数被约分掉，这就是可能的。但是因为分母和 p 是互素的 (p 是素数，而且 $(p-1)!$ 可以表示为小于 p 的数的乘积)，所以 p 的倍数不会被相约。哈哈！这就意味着我们只需证明上式中那个令人反感的分子可以被 p^2 整除就行了。这样做比取其他形式的分母要有效，因为现在我们只需证明关系式

$$\begin{aligned}&2\times 3\times\cdots\times(p-1)+1\times 3\times\cdots\times(p-1)+\cdots\\&+1\times 2\times 3\times\cdots\times(p-2)=0\ (\operatorname{mod} p^2)。\end{aligned}$$

(我们再次转到模算术上，这是通常证明一个数整除另一个数的最好的方法。然而，如果问题涉及多个整除性，例如要求被某个数的所有除数整除，有时其他技巧更有效。)

虽然我们得到了一个关系式，但它很复杂，需要进一步简化。关系式等号左边是一个不确定乘积的不确定和(这里“不确定”仅仅是指表达式中有“…”)，不过我们可以把这

个不确定乘积表示得更清楚些。设 i 是 1 和 $p-1$ 之间的一个数，每个不确定乘积都只是把从 1 到 $p-1$ 之间除 i 之外的所有数乘起来而得到的，所以可表达成更简洁的形式 $(p-1)!/i$。因为 i 和 p^2 是互素的，所以在模 p^2 下除以 i 是允许的[⑦]。于是我们的目标变为证明

$$\frac{(p-1)!}{1}+\frac{(p-1)!}{2}+\frac{(p-1)!}{3}+\cdots+\frac{(p-1)!}{p-1}=0 \pmod{p^2}。$$

我们对上式进行因式分解，得到

$$(p-1)!\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}\right)=0 \pmod{p^2}。\qquad (6)$$

(请记住，我们在进行模算术运算，所以像 1/2 这样的数是和某个整数相等的。例如，1/2=6/2=3 (mod 5)[⑦]。)

现在我们得到了以下形式的关系式：

$$(因子\ 1)\times(因子\ 2)=0 \pmod{p^2}。$$

如果不是在模算术中，我们可以说上式中的某个因子为零。在模算术中，我们几乎也可以得到同样的结论，但必须要慎重。幸运的是，第一个因子 $(p-1)!$ 和 p^2 是互素的(因为 $(p-1)!$ 和 p 是互素的)，故可以去掉。这样做的结果是式(6)等价于

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}=0 \pmod{p^2}。$$

(请注意，上式和我们原来的问题看起来非常类似，唯一的不同在于这里考虑的是整个分数，而不只是分子。没有细心的思考，我们不能从一种形式直接跳到另一种形式，所以上述复杂的推导是必要的。)

我们已经把问题转化为证明一个看起来更舒服的模算术关系式。但是接下来该做什么呢？也许举一个例子会有所帮助。让我们看看题目中给出的例子，即 $p=5$ 的情形。正如我



们所希望看到的，这时有

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=1+13+17+19 \pmod{25}=0 \pmod{25}。$$

但是上式成立的原因是什么呢？数字 1, 13, 17 和 19 看起来是随机的，但加起来却“神奇地”恰好是我们想要到的结论。也许这是巧合吧。让我们再尝试 $p=7$ 的情形：

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}=1+25+33+37+10+41 \pmod{49}$$
$$=0 \pmod{49}。$$

同样是“好运气”。怎么会这样呢？我们并不清楚所有的项是如何在模 p^2 下相互抵消的。牢记目标(b)，我们可以先证明(mod p)的情形，也就是说，先证明

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}=0 \pmod{p}。\tag{7}$$

即使得不到什么启发，也不失为一种有益的尝试。(而且，如果我们不能解决(mod p)的情形，那么根本就没有机会解决(mod p^2)的情形。)

事实上，证明关系式(7)要容易得多，因为它更简单。例如，当 $p=5$ 时，有

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=1+3+2+4 \pmod{5}$$
$$=0 \pmod{5};$$

而当 $p=7$ 时，有

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}=1+4+5+2+3+6 \pmod{7}$$
$$=1+2+3+4+5+6 \pmod{7}$$
$$=0 \pmod{7}。$$

现在，某种规律显现出来了：倒数 1/1, 1/2,⋯, 1/(p–1)恰好

覆盖所有的余数 1, 2,⋯, $(p-1)$ (mod p)一次[1]。例如，在以上 $p=7$ 时的等式中，1+4+5+2+3+6 可以重新排列为 1+2+3+4+5+6 的形式，在模 7 下这个数为 0。再验证 p 更大时的例子：在模 11 下可以得到

$$\begin{aligned}\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{10}&=1+6+4+3+9+2+8+7+5+10 \pmod{11}\\&=1+2+3+4+5+6+7+8+9+10 \pmod{11}\\&=0 \pmod{11}。\end{aligned}$$

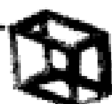
这表明，通过把倒数重排成这种有序的形式，可以很巧妙地解决模 p 的问题。然而这却不能轻易推广到模 p^2 的情形。与其拼命地把一块方积木塞进一个圆洞里(虽然用足够大的力气推，也能塞进去)，还不如找一块稍圆些的积木。因此我们要做的是找出证明 $1/1+1/2+\cdots+1/(p-1)=0$ (mod p)的另一种方法，希望它至少可以部分地推广到(mod p^2)的情形。

我们解决这类问题的经验现在该派上用场了。例如，通过问题 2.6 我们得知对称性或反对称性可能会有用，特别是在模算术中。在证明关系式(7)的过程中，通过用 -1 替换 $p-1$，用-2 替换 $p-2$，等等，我们得到

$$\begin{aligned}&\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{p-1}\\&\quad=\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{-3}+\frac{1}{-2}+\frac{1}{-1} \pmod{p}。\end{aligned}$$

然后可以很容易地配对并抵消(由于 p 是奇素数，所以不存在不成对的“中间项”)。我们能否在(mod p^2)的情形中进行同样的处理呢？答案是“有几分类似”。当求解(mod p)的问题时，我们把 $1/1$ 和 $1/(p-1)$，$1/2$ 和 $1/(p-2)$，等等，进行配对。当我们试图在(mod p^2)的情形中配对时，得到

$$\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{p-1}$$



$$=\left(\frac{1}{1}+\frac{1}{p-1}\right)+\left(\frac{1}{2}+\frac{1}{p-2}\right)+\cdots+\left(\frac{1}{(p-1)/2}+\frac{1}{(p+1)/2}\right)$$

$$=\frac{p}{1\times(p-1)}+\frac{p}{2\times(p-2)}+\cdots+\frac{p}{[(p-1)/2]\times[(p+1)/2]}$$

$$=p\left\{\frac{1}{1\times(p-1)}+\frac{1}{2\times(p-2)}\right.$$

$$\left.+\cdots+\frac{1}{[(p-1)/2]\times[(p+1)/2]}\right\}\ (\bmod p^2)。$$

乍一看，上式似乎变得更复杂而不是更简单了。但是我们在上式等号右边得到了一个非常重要的因子 p。因此，我们不必证明

$$(\text{表达式})=0\ (\bmod\ p^2),$$

而只需证明

$$(p\times\text{表达式})=0\ (\bmod\ p^2)。$$

这就等价于证明形如

$$(\text{表达式})=0\ (\bmod\ p)$$

的关系式。换句话说，我们把$(\bmod\ p^2)$的问题简化成$(\bmod\ p)$的问题，这就实现了前面给出的目标(b)。虽然这里表达式略显复杂了些，但问题被简化成模较小的情形还是很值得的。

由于$(\bmod\ p)$比$(\bmod\ p^2)$可以消去更多项，所以表达式看起来越来越复杂只是一种错觉。现在，我们只需证明

$$\frac{1}{1\times(p-1)}+\frac{1}{2\times(p-2)}+\cdots+\frac{1}{[(p-1)/2]\times[(p+1)/2]}$$
$$=0\ (\bmod p)。$$

但是 $p-1=-1\ (\bmod\ p)$，$p-2=-2\ (\bmod\ p)$，等等，所以上式简化为

$$\frac{1}{-1^2}+\frac{1}{-2^2}+\cdots+\frac{1}{-[(p-1)/2]^2}=0\ (\bmod p),$$

或等价于

$$\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{[(p-1)/2]^2}=0 \pmod p。$$

除了等号左边的最后一项有些令人意外(即止于 $1/[(p-1)/2]^2$，而不是如 $1/(p-1)^2$ 那样更自然的项)，上式还不算太糟。利用 $(-a)^2=a^2 \pmod p$这一事实，对上式“加倍”可以得到

$$\begin{aligned}&\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{[(p-1)/2]^2}\\&\quad=\frac{1}{2}\left\{\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{[(p-1)/2]^2}\right.\\&\qquad\left.+\frac{1}{(-1)^2}+\frac{1}{(-2)^2}+\cdots+\frac{1}{[-(p-1)/2]^2}\right\} \pmod p\\&\quad=\frac{1}{2}\left[\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{(p-1)^2}\right] \pmod p。\end{aligned}$$

所以证明 $1/1^2+1/2^2+\cdots+1/[(p-1)/2]^2=0 \pmod p$就等价于证明 $1/1^2+1/2^2+\cdots+1/(p-1)^2=0 \pmod p$。由于后者具有更好的对称性，所以处理起来更容易。(对于对称性，最好保留它，直到其作用得到充分发挥；而对于反对称性，消除得越早越好。)

所以为了证明原问题，我们只需证明

$$\frac{1}{1^2}+\frac{1}{2^2}+\cdots+\frac{1}{(p-1)^2}=0 \pmod p。\tag{8}$$

因原问题涉及分子和 p^2 的整除性(这是很强的性质，故很难处理)，而上式只涉及 p 的整除性，所以比原问题要好处理得多。

我们已经实现了所有的战术目标，并对问题做了适当简化。接下来该做什么呢？现在的问题看起来与我们前面考虑



的另一个关系式(7)紧密相关，但我们并没有原地绕圈。我们现在的目标——证明式(8)，蕴涵着原问题，而证明式(7)只是一个附带问题，要比原问题简单。我们希望朝着答案盘旋前进，而不是原地转圈。既然已经证明了式(7)，我们能否用同样的方法证明式(8)呢？

幸运的是，我们已经有了两种证明式(7)的方法：一种是把倒数重新排列的方法；另一种是配对抵消的方法。配对抵消法对式(7)有效，但对式(8)却行不通，主要原因是分母中的平方数会产生对称性，而不是反对称性。然而，倒数重排法对式(7)和(8)都有效可行。再以 $p=5$ 为例 (这样我们就可以利用某些前面的工作)：

$$\begin{aligned}\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}&=1^2+3^2+2^2+4^2 \pmod 5\\&=1^2+2^2+3^2+4^2 \pmod 5\\&=0 \pmod 5\text{。}\end{aligned}$$

当 $p=5$ 时可行表明，这个方法也可用于一般情形。从以上的例子可以看出，在模 p 下，全体剩余类[①] $1/1, 1/2, 1/3, \cdots, 1/(p-1)$ 恰好是全体数 $1, 2, 3, \cdots, p-1$ 的重新排列。这个事实的证明将会出现在讨论的最后。所以，我们可以说：在模 p 下，$1/1^2, 1/2^2, \cdots, 1/(p-1)^2$ 恰好是 $1^2, 2^2, 3^2, \cdots, (p-1)^2$ 的重新排列。换句话说，有

$$\begin{aligned}&\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{(p-1)^2}\\&=1^2+2^2+3^2+\cdots+(p-1)^2 \pmod p\text{。}\end{aligned}$$

上式消去了在求和中令人头痛的倒数，所以更容易处理了。事实上，利用标准公式(可以很容易地由归纳法证明)

$$1^2+2^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6},$$

我们可以求出这个和式，所以证明式(8)就转化为证明

$$\frac{(p-1)p(2p-1)}{6}=0 \pmod p。$$

当 p 是大于 3 的素数时，很容易证明上式成立(因为在这种情形中$(p-1)(2p-1)/6$是个整数)。

我们成功了。通过不断转化，关系式变得越来越简单，直到再也不能简化为止。这个过程有点儿冗长，但对于这些非常复杂的问题，逐步简化的方法有时是求解的唯一途径。

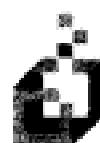
最后，我们来证明：在模 p 下，倒数 $1/1, 1/2, \cdots, 1/(p-1)$ 是 $1, 2, \cdots, (p-1)$的一个置换。这等价于说，模 p 的每个非零剩余都是唯一的一个模 p 的非零剩余数的倒数。这是很显然的[11]。

习题 2.5 设 n 是一个整数，$n\geqslant 2$。证明：

$$\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}$$

不是一个整数。(你需要使用 Bertrand 假设[12](实际上是一个定理)：对于任意给定的正整数 n，至少存在一个 n 和 $2n$ 之间的素数。)

习题 2.6 设 p 是一个素数，而 k 是一个不能被 $p-1$ 整除的正整数。证明：$1^k+2^k+3^k+\cdots+(p-1)^k$ 可以被 p 整除。(提示：因 k 可能是偶数，所以消去技巧不见得总能用得上；然而，重排技巧还是有效的。设 a 是 $\mathbf{Z}/p\mathbf{Z}$ 的一个生成元[20]，使得当 k 不是 $p-1$ 的倍数时，$a^k\neq 1 \pmod p$。然后用两种不同的方法计算表达式 $a^k+(2a)^k+\cdots+(p-1)^k$ 的值。)



第三章　代数和数学分析中的例子

我们不得不承认：这些数学公式是独立存在的，拥有它们自身的智慧……比起我们甚至那些发现者来，它们更加聪颖……我们从中获取的比为得到它们所付出的更多。

——Heinrich Hertz

大多数人都会把代数与数学联系起来。从某种意义来讲，这是有道理的。数学是研究抽象的对象，例如数值的、逻辑的或几何的对象，它们要满足一组精心选择的公理。初等代数研究的是满足上述数学定义的最简单且有意义的对象，它只有大约十几个假定，但却足以构成一个具有完美对称性的体系。作为一个例子，请看我偏爱的一个代数恒等式：

$$1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2\text{。}$$

这意味着前几位自然数的立方之和总是一个平方数，例如

$$1+8+27+64+125=225=15^2\text{。}$$

虽然有不止一种代数，但代数总是研究具有加、减、乘、除运算的数。数学中有很多种代数系统。例如，矩阵代数同样具有这四种运算，但其研究的对象是一组数，以它来代替一个数。其他的代数则使用其他各种运算和各种“数”。但令人惊奇的是，它们具有很多与初等代数相同的性质。例如，在某种特殊的条件下，方阵 $\boldsymbol{A}$ 满足代数方程

$$(\boldsymbol{I}-\boldsymbol{A})^{-1}=\boldsymbol{I}+\boldsymbol{A}+\boldsymbol{A}^2+\boldsymbol{A}^3+\cdots\text{。}$$

代数是大部分应用数学的基础。力学、经济学、化学、

电子学、最优化等领域的问题都可以用代数和微分学(这是代数的高等形式)来解决。事实上，代数是如此重要，以至于它的大部分奥秘都已经被揭开。所以它可以被放心地安排在高中课程中学习。然而，还是不时可以从中找到一些没被发现的小奥秘。

3.1 函数的分析

和代数学一样，分析学也是一门被广泛研究的学科。本质上，分析学是研究函数及其性质的；函数的性质越复杂，分析越“高深”。分析学的最低层次是研究满足简单代数性质的函数。例如，可以考虑一个函数 $f(x)$，满足：

$f(x)$是连续的，$f(0)=1$，且对于所有的实数 m, n，有

$$f(m+n+1)=f(m)+f(n), \tag{9}$$

然后推断这个函数的性质。在这个例子中，恰好存在一个满足条件的函数，即 $f(x)=1+x$。我们把它的证明留做练习。这类问题是学习如何用数学方法思考问题的一条很好的途径。因为题目中只有一两条信息可供使用，所以解题方向非常明确。这类问题可看做是一种“袖珍数学”，其中只有几个“公理”(即信息)可供使用，而不是有几十个公理和不计其数的定理需要考虑。然而，它同样会带给我们惊喜。

习题 3.1 设 f 是一个从实数到实数的函数，并满足条件(9)。证明：对于所有实数 x，有 $f(x)=1+x$。(提示：首先对整数 x 证明这个结论，然后讨论有理数 x，最后考虑实数 x。)

***问题 3.1** (Greitzer，1978，第 19 页) 假设 f 是一个把正整数映射到正整数的函数，使得对于所有正整数 n，满足 $f(n+1)>f(f(n))$。证明：对于所有正整数 n，$f(n)=n$。



这个问题看起来似乎没有给出足以证明结论的信息。毕竟，怎么可能用一个不等式来证明一个等式呢？其他此类问题(例如习题 3.1)一般都涉及函数**方程**。因为我们可以运用不同的代入法和类似的方法，逐渐把原始信息转换成一种便于处理的形式，所以处理起来更容易些。然而，问题 3.1 看起来完全不同。

如果仔细阅读题目，我们会注意到本题函数取的是整数值，而不像大多数问题那样常常是实数域上的函数方程。利用这一优势，可以立即构造一个“更强的”不等式：

$$f(n+1) \geqslant f(f(n))+1。 \tag{10}$$

现在让我们看看可以从中推导出什么结论。处理这类关系式的标准方法是给变量代入恰当的值。于是我们从 $n=1$ 入手：

$$f(2) \geqslant f(f(1))+1。$$

乍一看，这并没有告诉我们多少有关 $f(2)$ 或 $f(1)$ 的信息，但不等式右边的+1 提示我们 $f(2)$ 不可能太小。事实上，由于 f 映射到正整数，$f(f(1))$ 一定至少为 1，所以 $f(2)$ 至少为 2。需要证明的是 $f(2)=2$，因此也许我们的努力方向是对的。(尽量利用使你更能接近目标结论的策略。只有所有可能的直接方法都行不通，才考虑间接法，或者偶尔尝试回溯法。)

我们能否证明 $f(3)$ 至少等于 3 呢？再运用不等式(10)，得到 $f(3) \geqslant f(f(2))+1$。同样地，我们可以说 $f(3)$ 至少为 2。但是我们能否得出更强的结论呢？前面我们曾说过 $f(f(1))$ 至少为 1，也许可以证明 $f(f(2))$ 至少为 2。(实际上，因为我们“心中”知道 $f(n)$ 最终应等于 n，所以也就知道 $f(f(2))$ 应等于 2。但因为不可以利用要证的结果，所以我们现在不能用这个结论。)沿着这一思路，可以再通过不等式(10)得到

$$f(3) \geqslant f(f(2))+1 \geqslant f(f(f(2)-1))+1+1 \geqslant 3,$$

这里我们用 $f(2)-1$ 替代了不等式(10)中的 n。因为我们已经知道 $f(2)-1$ 至少等于 1，所以上式成立。

看起来我们可以推导出 $f(n) \geqslant n$ 了。因为我们用 $f(2)$至少为 2 的事实证明了 $f(3)$ 至少等于 3，所以正式的证明应利用归纳法。

这里运用归纳法需要一点技巧。考虑接下来的情形，即证明 $f(4) \geqslant 4$。由不等式(10)我们知道 $f(4) \geqslant f(f(3))+1$。由于 $f(3) \geqslant 3$，为了能够得到 $f(f(3))+1 \geqslant 4$，我们需要推出 $f(f(3)) \geqslant 3$。而要想证明它，就需要有结论"如果 $n \geqslant 3$，则 $f(n) \geqslant 3$"。证明这一类结论最便捷的途径是把它包含在我们要证明的归纳结论中。更确切地说，我们将证明：

引理 3.1 对于所有的 $m \geqslant n$，有 $f(m) \geqslant n$。

证明 我们对 n 运用归纳法。

- 基础情形($n=1$)：已知 $f(m)$是一个正整数，所以 $f(m)$ 至少为 1。故结论显然成立。

- 归纳递推：假设引理对 n 成立，于是我们试图证明对于所有 $m \geqslant n+1$ 有 $f(m) \geqslant n+1$。对于任意的 $m \geqslant n+1$，由不等式(10)得到 $f(m) \geqslant f(f(m-1))+1$。由于 $m-1 \geqslant n$，所以 $f(m-1) \geqslant n$(由归纳假设得出)。进而，既然 $f(m-1) \geqslant n$，那么再由归纳假设得到 $f(f(m-1)) \geqslant n$。因此 $f(m) \geqslant f(f(m-1))+1 \geqslant n+1$。由归纳法知引理 3.1 成立。 □

考虑引理 3.1 的特殊情形 $m=n$，就得到子命题：对于所有正整数 n，有

$$f(n) \geqslant n。 \tag{11}$$

接下来该做什么呢？正如所有关于函数方程的问题一样，一



旦有了一个新结果，就应该应用一番，尝试把它与前面的结果结合起来。之前我们唯一的结果是不等式(10)，所以可以把新的结果代入其中。在不等式(11)中用 $f(n)$代替 n，就可以得到下面有用的结果：

$$f(n+1)\geqslant f(f(n))+1\geqslant f(n)+1,$$

即

$$f(n+1)>f(n)。$$

这是一个很有用的关系，它意味着 f 是一个严格递增函数！(但从不等式(10)看，这并不是显而易见的吧？)于是得到 $f(m)>f(n)$当且仅当 $m>n$。这说明我们原来的不等式

$$f(n+1)>f(f(n))$$

可以被表达为

$$n+1>f(n)。$$

由此及不等式(11)，就证明了我们的结论。

问题 3.2 (Australian Mathematics Competition，1984，第 7 页) 假设 f 是一个定义在全体正整数上取整数值的函数，并具有性质：

(a) $f(2)=2$；

(b) 对于所有正整数 m 和 n，有 $f(mn)=f(m)f(n)$；

(c) 如果 $m>n$，则 $f(m)>f(n)$。

求 $f(1983)$ 的值，并给出理由。

我们需要找出 f 的一个特定值。最好的方法是设法推算 f 所有的值，而不仅仅是 $f(1983)$。(1983 只是问题提出的年份而已。) 当然，这里假定了 f 仅有一个解。但是，这个问题暗示了这样一个事实：$f(1983)$ 只有一个可能值(否则答案就不唯一了，再因为 1983 是一个很普通的数，我们有理由猜测 f

只有一个解。)

那么，f 有什么性质呢？我们知道 $f(2)=2$，反复应用性质(b)，得到 $f(4)=f(2)f(2)=4$, $f(8)=f(4)f(2)=8$，等等。实际上，简单的归纳可以证明，对于所有的 n，有 $f(2^n)=2^n$。所以当 x 是 2 的幂时，$f(x)=x$。也许 $f(x)=x$ 对于所有的 x 都成立。对 $f(x)=x$ 验证性质(a), (b)和(c)，就可以看出 $f(x)=x$ 是满足这三条性质的一个解。所以，如果我们认为只有一个这样的解 f，那么就一定是它了。因此，我们可以证明一个更普遍且更清晰的命题：

> 从正整数到整数并满足性质(a), (b)和(c)的唯一函数是恒等函数(也就是说，对于所有的 n，有 $f(n)=n$)。

因此我们需要证明：如果 f 满足性质(a), (b)和(c)，则 $f(1)=1$，$f(2)=2$，$f(3)=3$，等等。首先我们来证明 $f(1)=1$(对于函数方程，我们应先尝试一些小的值，从而对这个问题找到一点“感觉”)。由性质(c)我们知道 $f(1)<f(2)$，而且我们还知道 $f(2)=2$，所以 $f(1)<2$。由性质(b)(取 $n=1,m=2$)，我们得到

$$f(2)=f(1)f(2),$$

从而有

$$2=2f(1)。$$

这意味着 $f(1)$一定等于 1，正如我们想要的一样。

现在，我们已经有了 $f(1)=1$ 和 $f(2)=2$，那么 $f(3)$呢？性质(a)无济于事，而性质(b)只意味着 $f(3)$ 可以用来表示 $f(6)$ 或 $f(9)$ 等其他数，也没有太大的帮助。由性质(c)得到

$$f(2)<f(3)<f(4),$$



又 $f(2)=2$，$f(4)=4$，故

$$2<f(3)<4。$$

但 2 和 4 之间的唯一整数为 3，所以 $f(3)$ 一定是 3。

这给我们提供了一条线索：$f(3)$ 等于 3 只因为 $f(3)$ 是一个整数。(这与问题 3.1 中的 $f(n+1)>f(f(n))$ 类似。看出怎么回事儿了吗？)如果没有这一限制，$f(3)$ 可以是 2.1，3.5 或任何其他值。现在让我们看看能否更大程度地利用这一线索。

我们已经知道 $f(4)=4$，下一步求解 $f(5)$。我们希望用处理 $f(3)$ 的方法来确定 $f(5)$。由性质(c)得到

$$f(4)<f(5)<f(6)。$$

而 $f(4)=4$，但 $f(6)$ 呢？不用担心，6 等于 2 乘 3，所以 $f(6)=f(2)f(3)=2\times3=6$。由于 $f(5)$ 是 4 和 6 之间的整数，故必定为 5。这种策略看起来很顺利，我们已经确定了直到 $n=6$ 的 $f(n)$ 的所有值。

因为我们是依靠旧结果得到新结论的，所以对一般性的证明感觉上有很强的归纳法味道。又因为我们需要的不仅仅是前一个结果，而是前面的若干个结果，所以可能需要使用所谓的强归纳法(strong induction)[14]。

经分析，要解决问题 3.2，需先证明以下命题：

引理 3.2　对于所有的正整数 n，有 $f(n)=n$。

证明　我们运用强归纳法。首先验证初始情形：$f(1)=1$ 吗？是的，我们已经证明过这一结论。接下来假设 $m\geq2$，且对于所有小于 m 的正整数 n，有 $f(n)=n$。我们想要证明 $f(m)=m$。观察若干例子后不难发现，我们需要把问题分成两种情形来考虑：m 是偶数和 m 是奇数。

情形 1：m 是偶数。这时，我们可以把 m 记为 $m=2n$（n 为正整数），n 小于 m。由强归纳假设得到 $f(n)=n$，所以 $f(m)=f(2n)=f(2)f(n)=2n=m$。这正是我们想要证明的。

情形 2：m 是奇数。这时，我们可以把 m 写成 $m=2n+1$。由性质(c)得到 $f(2n)<f(m)<f(2n+2)$。因为 $n+1$ 和 $2n$ 都小于 m，所以由强归纳假设得到 $f(2n)=2n$ 和 $f(n+1)=n+1$。再由性质(b)得到 $f(2n+2)=f(2)f(n+1)=2(n+1)=2n+2$，故我们的不等式变为

$$2n<f(m)<2n+2,$$

从而得到 $f(m)=2n+1=m$。这正是我们想要的结果。

因此，无论在哪种情形中，$f(m)=m$ 都是成立的。于是，利用强归纳法我们证明了 $f(n)$ 等于 n。 □

所以问题 3.2 的答案一定是 $f(1983)=1983$，太棒了！

习题 3.2 如果我们用更弱的条件“(a') 对于至少一个整数 $n\geqslant 2$，有 $f(n)=n$”来替代(a)，证明问题 3.2 仍然可以求解。

***习题 3.3** 如果我们允许 $f(n)$ 取实数，而不只是整数，证明问题 3.2 仍然可以求解。(提示：首先通过对于不同的整数 n, m 比较 $f(2^n)$和 $f(3^m)$来证明 $f(3)=3$。)作为一个进一步的挑战，利用这个假设并用性质(a')替代性质(a)，求解问题 3.2。

****习题 3.4** (国际数学奥林匹克，1986，第 5 题) 设 f 为把非负实数映射到非负实数的函数，求所有满足下列条件的 f(如果存在的话)：



(a) 对于所有的非负实数 x, y，有

$$f(xf(y))\,f(y)=f(x+y);$$

(b) $f(2)=0$;

(c) $f(x)\neq 0$，当 $0\leqslant x<2$ 时。

(提示：第一个条件涉及函数值的乘积，而其他两个条件涉及函数取零值(或非零值)。那么当函数值的乘积等于零时，可以得到什么结论呢？)

3.2　多项式

很多代数问题与含有一个或多个变量的多项式有关，所以让我们先回顾关于这类多项式的若干定义和结论。

一元多项式，记做 $f(x)$，是一个具有以下形式的函数：

$$f(x)=a_nx^n+a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+\cdots+a_1x+a_0,$$

或者利用求和号记做

$$f(x)=\sum_{i=0}^{n}a_ix^i,$$

这里 $a_i(i=1,2,\cdots,n)$ 是常数(在本书中我们总是假定它为实数)且假定 a_n 不为零。我们称 n 为多项式 f 的**次数**。

多元多项式没有一元多项式那样漂亮的形式，但还是十分有用的。作为例子，假定有三个变量，如果 $f(x, y, z)$具有以下形式：

$$f(x,y,z)=\sum_{k,l,m}a_{k,l,m}x^ky^lz^m,$$

就称它是一个三元多项式，这里 $a_{k,l,m}$ 是(实)常数；求和号表示对所有满足 $k+l+m\leqslant n$ 的非负整数 k，l 和 m 求和，并且假定至少有一个非零的 $a_{k,l,m}$ 满足 $k+l+m=n$。n 被称做 f 的

次数。次数为 2 的多项式称为**二次多项式**，次数为 3 的多项式称为**三次多项式**，依此类推。如果次数为 0，就称多项式是**平凡的**或是**常数**。如果所有非零的 $a_{k,l,m}$ 都满足 $n=k+l+m$，就称多项式是**齐次的**。对于所有的 $x_1,x_2,\cdots,x_m$ 和 t，齐次多项式 f 满足

$$f(tx_1,tx_2,\cdots,tx_m)=t^m f(x_1,x_2,\cdots,x_m)。$$

例如，x^2y+z^3+xy 是含有三个变量(x, y 和 z)的多项式，且它的次数为 3。因为 xy 项的次数为 2，所以它不是齐次的。

如果一个含有 m 个变量的多项式 f 对于所有的 $x_1,x_2,\cdots,x_m$ 都满足 $f(x_1,x_2,\cdots,x_m)=p(x_1,x_2,\cdots,x_m)q(x_1,x_2,\cdots,x_m)$，就称 f **被因式分解**为两个多项式 p 和 q 的乘积，其中 p 和 q 称为 f 的因式。很容易证明，一个多项式的次数等于其因式的次数之和。如果一个多项式不能分解为非平凡的因式乘积，就称它是**不可约的**。

使得 $f(x_1,x_2,\cdots,x_m)=0$ 的一组值$(x_1,x_2,\cdots,x_m)$称为多项式 $f(x_1,x_2,\cdots,x_m)$的**根**。一元多项式的根的个数可以与多项式的次数一样多。事实上，如果把重根和复根计算在内，一元多项式的根的个数总是恰好等于多项式的次数。例如，二次多项式 $f(x)=ax^2+bx+c$ 的根可以由众所周知的二次方程求根公式

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

给出。对于三次和四次多项式，也有关于根的公式，但它们复杂得多，在实际问题中不是很有用。对于五次和更高次的多项式，就不存在根的初等公式了！而含有两个或更多个变量的多项式甚至更糟，这样的多项式通常有无穷多个根。



一个因式的所有的根是原多项式的所有的根的一个子集。在判定一个多项式能否整除另一个多项式时这是非常有用的信息。特别地，因为 a 是 $x-a$ 的根，所以 $f(x)$可以被 $x-a$ 整除当且仅当 $f(a)=0$。而且，对于任意的一元多项式 $f(x)$和任意实数 t, $x-t$ 整除 $f(x)-f(t)$。

让我们考虑若干与多项式有关的例题。

问题 3.3(Australian Mathematics Competition, 1987，第 13 页） 设 a, b, c 是满足

$$\frac{1}{a}+\frac{1}{b}+\frac{1}{c}=\frac{1}{a+b+c} \tag{12}$$

的实数，上式中的所有分母都不等于零。证明

$$\frac{1}{a^5}+\frac{1}{b^5}+\frac{1}{c^5}=\frac{1}{(a+b+c)^5}. \tag{13}$$

这个问题乍看起来很简单。题目只给定了一个条件，所以应该可以通过一步步的逻辑推导来直接证明我们的结论。想从式(12)推出式(13)的一个最初的想法可能是要把式(12)的两边都上升到 5 次幂，使它与所要证明的结果更接近，但这样做导致等号左边的项变得非常复杂。看起来也没有其他显而易见的处理方法，所以直接推导方法只能到此为止。

再观察一下，会发现式(12)很有欺骗性，它很像高中生常被告诫不要轻易利用的那类关系式，因为它们通常会使人误入歧途。这就给我们提供了第一条有用的线索：式(12)应对 a, b, c 的取值作更一进步的约束。所以，重新考虑式(12)可能是必要的。

取公分母应是一个良好的开端。把等号左边的三个倒数相加，我们得到

$$\frac{ab+bc+ca}{abc}=\frac{1}{a+b+c},$$

再交叉相乘得到

$$ab^2+a^2b+a^2c+ac^2+b^2c+bc^2+3abc=abc。\qquad (14)$$

到了这一步有人可能会想到以下各种不等式会在这里派上用场，如柯西-施瓦茨(Cauchy-Schwarz)不等式，算术平均值-几何平均值不等式，等等(Hardy，1975，第33～34页)。如果 a, b, c 被限制为正的，也许有所帮助，但题目中并没有给出这样的约束。事实上，这个条件不能成立，因为如果 a, b 和 c 都为正数，那么 $1/(a+b+c)$ 比等式(12)等号左边的所有三个倒数都要小。

因为式(14)等价于式(12)，且在代数意义上式(14)更简单(式(14)不包含倒数)，所以我们可以尝试从式(14)来推出式(13)。在这种情况下，直接方法也是不可行的。要想从若干个关系式推出另一个关系式，通常其他唯一的途径是证明一个中间结果或作某种有用的变量替换。(还有一些非寻常的方法，例如，把式(12)看做是函数 $1/a+1/b+1/c-1/(a+b+c)$ 的等值线，然后利用微积分知识来得到这条等值线的形状和性质，但最好还是坚持尝试一些初等方法)。

变量替换法看起来不是理想的选择，因为式(12)或(14)本身已足够简洁，变量替换法很难使它们变得更简洁。因此我们将设法来猜测并证明一个中间结果。这个中间结果最好是具有参数化形式，因为这样可以直接把参数代入所要证明的关系式中。进行参数化的一种方法是解出其中一个变量，例如 a。从式(14)中解出 a 并不容易(除非你愿意使用二次方程求根公式)，而从(12)式中确实可以解出 a。通过依次解出 a, b 和 c 并推出一个中间结果，就可以证明我们的问题。(这个中间结果恰好等价于我们将在下面给出的结果。它就该是



如此，不是吗？)但是这里我将尝试其他解法。

放弃参数化，可以很方便地把式(14)改写成一种更好的形式。满足式(14)的量本质上就是多项式[⑤]

$$ab^2+a^2b+a^2c+ac^2+b^2c+bc^2+2abc \tag{14'}$$

的根。处理多项式的根的最佳方法是对多项式进行因式分解(反之亦然)。那么它的因式是什么呢？因为我们知道式(14)必须以某种方式来推出式(13)，所以应该确信一定存在式(14)的某种便于运算的形式，用它可以导出式(13)[●]，而一个多项式的唯一便于运算的形式是对它作因式分解。为了找出因式，我们需要尝试。因多项式(14′)是齐次的，故因式也应是齐次的；这个多项式是对称的，故这些因式也应是相互对称的；这个多项式是三次的，故应存在一个线性因式。于是我们应检验形如 $a+b$, $a-b$, a, $a+b+c$, $a-b+c$ 的因式($a+2b$ 之类的也可能成立，但不太简洁，所以可以留到后面再尝试)。这样很快可以看出(从因式定理)，$a+b$ 是该三次多项式的因式；类似地还可以得到 $b+c$ 和 $c+a$ 也是该三次多项式的因式。由此很容易验证多项式(14′)可以因式分解为 $(a+b)(b+c)(c+a)$。这意味着，式(14)，即式(12)成立当且仅当 $a+b=0$ 或 $b+c=0$ 或 $c+a=0$。将每一种可能的解代入(13)，即可证明所要的结果。

习题 3.5 因式分解 $a^3+b^3+c^3-3abc$。

习题 3.6 找出满足 $a+b+c+d=0$ 和 $a^3+b^3+c^3+d^3=24$ 的所有整数 a, b, c, d。(提示：这两个方程的某些解并不难猜出来，但为了证明你找到了所有解，需要把第一个方程代入第二个方程中，然后进行因式分解。)

● 译者注：原文为 $5ab$，应为笔误，译者改为(13)。

多项式的可分解或不可分解性是数学中一个充满魅力的课题。接下来的问题很有启发性，因为对它的解答几乎用到了本书中的所有技巧。

> **问题 3.4 证明：任意具有如下形式的多项式
> $$f(x)=(x-a_1)^2\cdots(x-a_n)^2+1$$
> 不能分解成两个整系数的非平凡多项式的乘积，其中 $a_1, \cdots, a_n$ 是两两不同的整数[⑧]。**

这是一个相当一般性的命题。例如，它表明多项式

$$(x-1)^2(x+2)^2+1=x^4+2x^3-3x^2-4x+5$$

不能分解出其他整系数的多项式因式。我们该如何证明这个命题呢？

假若 $f(x)$ 可分解成两个整系数的非平凡多项式 $p(x)$ 和 $q(x)$ 的乘积，那么对于所有的 x 有 $f(x)=p(x)q(x)$。这可是个重要的信息。但是请记住，f 具有一种特性：它是某个平方式加 1。我们该如何利用这一性质呢？我们可以说 $f(x)$ 总是正的(或者甚至可以说 $f(x)\geqslant 1$)，但这除了说明 $p(x)$ 和 $q(x)$ 符号相同外，并没有提供太多的信息。然而，我们还有另一条信息：f 并不只是任意通常意义的平方式加 1，其中的平方式是若干个线性因式 $(x-a_i)$ 的乘积取平方。我们能否利用这些对我们有利的 $(x-a_i)$ 呢？

利用因式的最好方法是使它等于 0，因为这样可以使整个表达式为 0。(当然，有时因为希望消去某个因式，那么这个因式等于 0 就是最不希望看到的了。)当 $x=a_i$ 时，$x-a_i=0$，于是我们想到用 a_i 替代 x，从而得到

$$f(a_i)=\cdots(a_i-a_i)^2\cdots+1=1。$$



回到 $p(x)$和 $q(x)$，上式就意味着

$$p(a_i)q(a_i)=1。$$

而这又意味着什么呢？如果我们忘记 p 和 q 具有整系数，也不记得 a_i是整数，那它就几乎没什么意义了。上式的关键是 $p(a_i)$和 $q(a_i)$都是整数。因此我们得到了两个整数相乘等于 1 的结论。这只可能在两个整数同时为+1 或–1 时发生。简而言之，对于所有的 $i=1, \dots, n$，有

$$p(a_i)=q(a_i)=\pm 1。$$

我们应该注意这里的符号“±”。在现在我们仅知道，例如 $p(a_1)$ 和 $q(a_1)$相等，但 $p(a_1)$和 $p(a_2)$的符号也许相同，也许相反。

我们已经几乎确定了 $p(a_1), \cdots, p(a_n)$ 和 $q(a_1), \cdots, q(a_n)$ 的值，因此多项式 p 和 q 都在 n 个点上被“固定”了。但是，多项式仅有与它的次数相等的自由度。因为 $pq=f$，所以 p 的次数加 q 的次数等于 f 的次数 $2n$。这意味着其中一个多项式，设为 p，它的次数最多为 n。总之，我们得到了一个次数最多为 n 但在 n 个点上已被约束的多项式。看来有希望利用这些事实来导出矛盾，这就是我们要进一步研究的。

对于一个次数最多为 n 的多项式，我们了解些什么？它最多有 n 个根。对于 p 的根，我们又了解些什么？p 是 f 的因式，所以 p 的根也是 f 的根。f 的根是什么？它根本没有根(至少在实数轴上没有根)！这是因为 f 总是正的(事实上，它总是至少为 1)，所以没有根。这就意味着 p 也没有根。一个多项式没有根的几何意义是什么呢？它意味着多项式的图像不会穿过 x 轴，也就是说，它不会改变符号。换句话说，p 总是正的或者总是负的。这样我们只需考虑两种情形。但如果注意到由其中一种情形可推出另一种情形，就可以省些力气。实际上，如果我们有一种因式分解 $f(x)=p(x)q(x)$，自然

就有了另一种因式分解：$f(x)=(-p(x))(-q(x))$。所以，如果 p 总是负的，我们总是可以对这一分解式的因式都取负号，得到一种新的分解，使得 p 总是正的。

因此，不失一般性，我们可以假定 p 总是正的。我们已经知道 $p(a_i)=+1$ 或-1，现在又知道它还是正的，所以对于所有的 i，$p(a_i)$一定为$+1$。而且，因为 $q(a_i)$必定等于 $p(a_i)$，所以对于所有的 i，$q(a_i)$也为$+1$。接下来做什么呢？

$p(x)$和 $q(x)$都必定要取至少 n 次$+1$ 的值，这一结论可以从根的角度来重新叙述如下：$p(x)-1$ 和 $q(x)-1$ 都至少有 n 个根。但是因为 $p(x)$的次数最多为 n，所以 $p(x)-1$ 的次数最多也是 n。这意味着只有当 $p(x)-1$ 的次数恰好为 n 时，$p(x)-1$ 才可能有 n 个根[⑧]。于是 $p(x)$的次数为 n，从而 $q(x)$的次数也是 n。

小结一下到目前为止我们所掌握的信息：我们假设 $f(x)=p(x)q(x)$；p 和 q 都是次数为 n 且取值为正的整系数多项式，且对于所有的 $i(1\leqslant i\leqslant n)$，有 $p(a_i)=q(a_i)=1$，即 $p(a_i)-1=q(a_i)-1=0$。于是我们知道 $p(x)-1$ 的根就是 a_i。因为 $p(x)-1$ 最多只可能有 n 个根，所以这 n 个两两不同的 a_i[⑨]是 $p(x)-1$ 仅有的根。这意味着 $p(x)-1$ 具有以下形式：

$$p(x)-1=r(x-a_1)(x-a_2)\cdots(x-a_n);$$

类似地，$q(x)-1$ 具有以下形式：

$$q(x)-1=s(x-a_1)(x-a_2)\cdots(x-a_n),$$

这里 r 和 s 是某两个常数。为了了解更多有关 r 和 s 信息，请记住 p 和 q 都是整系数多项式。$p(x)-1$ 的首项系数是 r，$q(x)-1$ 的首项系数是 s，这意味着 r 和 s 一定是整数。

现在我们把 $p(x)$和 $q(x)$的这些表达式代入原有的表达式 $f(x)=p(x)q(x)$中，得到



$$(x-a_1)^2(x-a_2)^2\cdots(x-a_n)^2+1$$
$$=(r(x-a_1)(x-a_2)\cdots(x-a_n)+1)$$
$$\times(s(x-a_1)(x-a_2)\cdots(x-a_n)+1)。$$

上式联系着两个明确给出的多项式，接下来应做的最好的事是比较系数。

比较 x^{2n} 的系数我们得到 $1=rs$。因为 r 和 s 都是整数，所以有 $r=s=+1$ 或 $r=s=-1$。让我们先假设 $r=s=1$。上面的多项式等式变为

$$(x-a_1)^2(x-a_2)^2\cdots(x-a_n)^2+1$$
$$=((x-a_1)(x-a_2)\cdots(x-a_n)+1)$$
$$\times((x-a_1)(x-a_2)\cdots(x-a_n)+1)。$$

通过展开和消去处理，上式变为

$$2(x-a_1)(x-a_2)\cdots(x-a_n)=0。$$

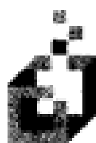
这是十分荒谬的结论(因上式必须对所有的 x 成立)。对 $r=s=-1$ 时的讨论是类似的。证毕。

习题 3.7　证明：多项式

$$f(x)=(x-a_1)(x-a_2)\cdots(x-a_n)+1$$

不能分解成两个次数较低的整系数多项式的乘积，其中 a_i 是两两不同的整数[⑰]。(提示：假设 $f(x)$ 可分解为两个多项式 $p(x)$ 和 $q(x)$ 的乘积，考虑 $p(x)-q(x)$。请注意，这种特殊的策略也可用于问题 3.4，但事实上不是很有效。)

习题 3.8　设 $f(x)$ 是一个整系数多项式，且 a, b 是整数。证明：仅当 a, b 是相邻的整数时，$f(a)-f(b)$ 才可能等于 1。(提示：对 $f(a)-f(b)$ 进行因式分解。)



第四章　欧几里得几何

当爱斯奇里斯●(Aeschylus)被人们遗忘时，阿基米德(Archimedes)会依然被铭记，这是因为语言可以失去生命力，而数学思想却永葆青春。

G. H. Hardy，《一个数学家的辩白》[⑱]

欧几里得几何学可以看成是数学中第一个具有现代风格(用到假设、定义、定理等)的分支。即使到现在，欧几里得几何学仍保持着逻辑性强且结构严密的传统。它的若干基本结果可以用来系统地处理并解决与几何对象和几何思想有关的问题。这些思想在与解析几何相结合时被发挥得淋漓尽致。解析几何把点、线、三角形和圆放到被划分为四部分的平面坐标系中，从而把几何问题很自然地转化为代数问题。但是几何的真正的美在于它可以通过反复运用显而易见的事实，来准确无误地证明看起来并不很明显的结论。这里，以泰勒斯(Thales)定理(Euclid III, 31)[⑲]为例作说明。

定理 4.1 (泰勒斯定理)　圆的直径所对应的圆周角是直角。换句话说，在图 1 中，有$\angle APB=90^\circ$。

证明　设 O 为圆心。连接线段 OP，就把 $\triangle APB$ 分成了两个等腰三角形(因为$|OP|=|OA|$和$|OP|=|OB|$，这里我们用 $|AB|$ 表示线段 AB 的长度)。利用“等腰三角形的底角相等”及“三角形的内角和为 180°”，得到

● 译者注：爱斯奇里斯(公元前 525 年—456 年)是古希腊著名悲剧作家，一生创作了约 70 出剧作。他被公认为“希腊悲剧之父”。



$$\angle APB = \angle APO + \angle OPB = \angle PAO + \angle PBO$$
$$= \angle PAB + \angle PBA = 180^\circ - \angle APB,$$

所以 $\angle APB$ 一定是直角。　□

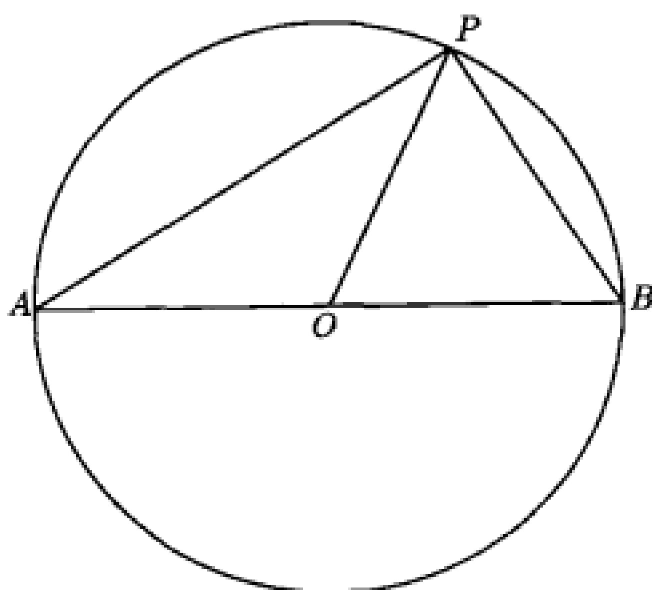


图　1

几何中经常出现这样的情况：一些结论你可以通过画图、测量角度或长度来验证，但却不能立即给出证明。例如定理：一个四边形的四条边的中点总是构成一个平行四边形。像这些结论肯定具有它本身的某些内在结构。

> **问题 4.1** (Australian Mathematics Competition，1987，第 12 页)　设 $\triangle ABC$ 是圆的内接三角形，其三个内角 $\angle A, \angle B, \angle C$ 的角平分线分别交圆于点 D, E, F。证明：AD 垂直于 EF。

证明的第一步当然是画张示意图(图 2)，并把已知信息尽量标出。这里我们也标出了三角形的内心 I (三条角平分线的交点，它会是很重要的)以及 AD 与 EF 的交点 M(就是我们想要证明是直角的地方)。这样，我们就可以将需要证明的目标写成一个等式：$\angle AMF = 90^\circ$。

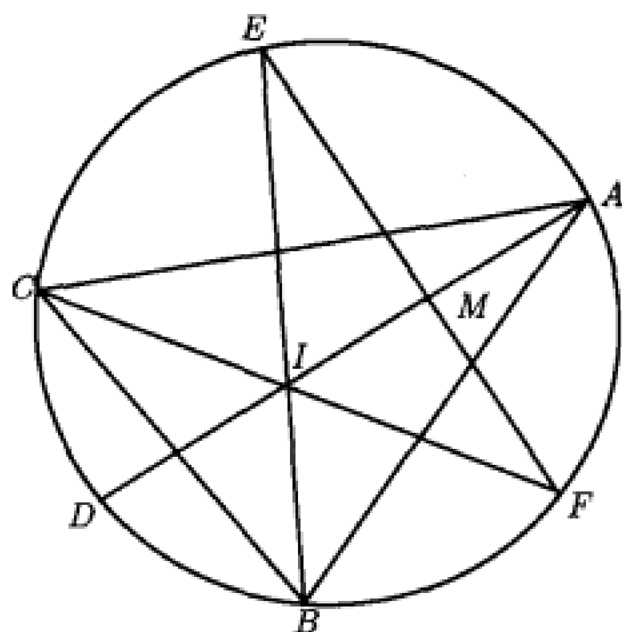


图 2

这是一个很直观的问题，示意图很容易画，结论在图中也十分明显。对于这样的问题，直接方法可能就行得通。

我们需要计算在点 M 处的角。乍一看，M 点并没有什么特别。但是在补充某些信息后就会发现我们已经获得了有关其他角的丰富信息，这主要归功于各个角平分线以及三角形和圆。也许只要求得足够多的角，我们就可能求出$\angle AMF$。另外，我们有许多定理可以利用：三角形的内角和为 180°；一条弧上的弦所对应的圆周角相等；三条内角平分线共点；等等。

我们需要从某些角入手。注意到“主”三角形是$\triangle ABC$，而所有有关角平分线、圆及其他信息都是围绕着这个三角形的，所以最好可能是从这个三角形的角$\alpha=\angle BAC$, $\beta=\angle ABC$, $\gamma=\angle BCA$(习惯上用希腊字母来标记角)开始讨论。当然，我们有$\alpha+\beta+\gamma=180°$。我们还可以添加许多其他角的信息，例如$\angle CAD=\alpha/2$。(最好是你自己动手画一张草图并标出角度。)这样，我们就可以利用三角形的内角和为 180° 这一事实，计算出其他内角。例如，如果 I 是$\triangle ABC$ 的内心(AD, BE 和



CF 的交点)，考虑△AIC 就可很容易地得到$\angle AIC=180°-\alpha/2-\gamma/2$。事实上，除了那些在点 M 处的角外，我们几乎可以得到所有相关的角。而在点 M 处的角才是我们真正需要的。因此，我们必须用与点 M 无关的角来表示在点 M 处的角。这很容易做到。例如，可以把我们期望等于 90° 的$\angle IMF$ 写成

$$\angle IMF = 180° - \angle MIF - \angle IFM = 180° - \angle AIF - \angle CFE.$$

再因为$\angle AIF$ 和$\angle CFE$ 比$\angle IMF$ 容易计算得多，所以上式进了一步。的确，我们有

$$\angle AIF = 180° - \angle AIC = \alpha/2 + \gamma/2。$$

再因为等长的弦所对应的圆周角相等，所以我们又有

$$\angle CFE = \angle CBE = \beta/2。$$

因此

$$\angle IMF = 180° - \alpha/2 - \beta/2 - \gamma/2 = 180° - 180°/2 = 90°。$$

这就是我们想要证明的结论。

直接计算角度是很适合求解某些几何问题的一种方法。角通常比边更容易计算(计算边要处理各种繁琐的正弦、余弦定理)，而且计算法则也更容易记忆。对于那些与边长无关，但与许多三角形和圆(尤其是等腰三角形)有关的问题，这种方法最为有效。但是对于一些比较难求的角，你通常需要先计算出许多其他的角。

> **问题 4.2** (Taylor, 1989, 第 8 页, 问题 1) 在△BAC 中，$\angle B$ 的角平分线交 AC 于点 D，$\angle C$ 的角平分线交 AB 于点 E。这两条角平分线相交于点 O。假设 $|OD|=|OE|$。证明：$\angle BAC=60°$或△BAC 为等腰三角形(两个结论可以同时成立)。

我们首先应画张示意图。因为必须使 OD 和 OE 等长，所以画起来需要一点儿技巧，但是我们可以略施小计，使 $\triangle ABC$ 成为等腰三角形或 $\angle BAC=60°$(既然我们知道这是一定成立的)。这样就产生了两种可能的画法(图 3)。

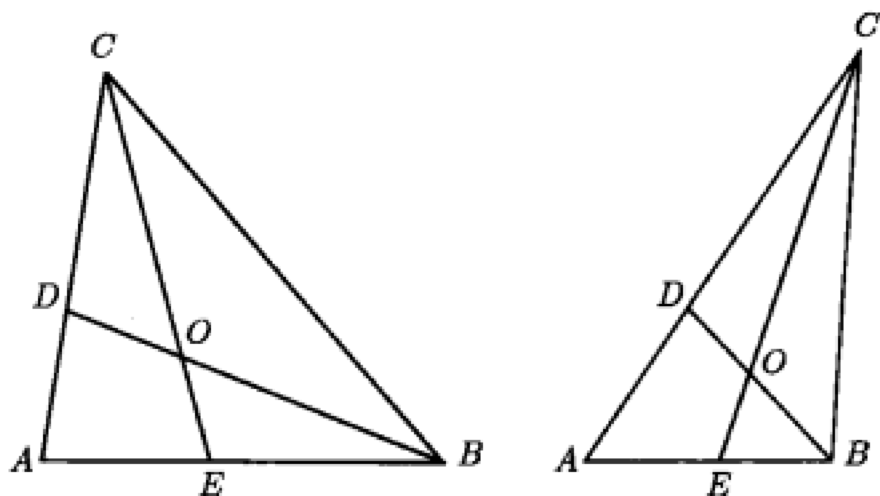


图 3

我们只有一个条件：$|OD|=|OE|$，而希望证明一个看起来不同寻常的结论：关于这个三角形的两种性质之一成立。但是这两种性质都与角相关(等腰三角形的底角相等，而角平分线显然与角有关)，因此本题可以看做是一个有关角的问题(至少一开始可以这样认为)。

一旦我们确定了问题是与角相关的，接下来要做的就是把给定的条件 $|OD|=|OE|$ 用角重新表示出来。因为 $\triangle ODE$ 是等腰三角形，所以显然 $\angle ODE=\angle OED$。这使我们看到了一丝希望，但还没有把 $\angle ODE$ 和 $\angle OED$ 与其他角联系起来。特别是，我们想要把这两个角用 $\alpha=\angle BAC$, $\beta=\angle ABC$, $\gamma=\angle ACB$ 表示，这是因为我们想要证明的是 $\alpha=60°$ 或 $\beta=\gamma$。(而且，$\triangle ABC$ 是“主”三角形，所有其他信息都是从这个三角形衍生出来的。它是一个逻辑上的参考系，所有量都应利用这个“主”三角形来表示。)不过也存在其他将边转化为角的方法。



让我们看看 OD 和 OE。我们想要把这两条边的长度与角 α, β, γ 联系起来。联系边与角的方法有几种：利用初等三角学，相似三角形，等腰和等边三角形，正弦和余弦定理等，这里提到的只是其中的一部分。其中，初等三角学需要用到直角和圆，但我们并没有很多这方面的信息。我们也没有什么相似三角形，等腰三角形法又已经考虑过了。余弦定理通常使问题复杂化而不是使之变得简单，因为它涉及更多未知的边长。这样就只剩下正弦定理可以选择了。毕竟，它可以将边与角直接联系起来。

为了使用正弦定理，我们需要一个或两个三角形，最好是那些包含 OD, OE 并含有多个已知角的三角形。通过观察示意图并估测角度，我们可以猜想$\triangle AOD$, $\triangle COD$, $\triangle AOE$, $\triangle BOE$ 可能派上用场。$\triangle AOE$ 和$\triangle AOD$ 有一条公共边，这应该使问题变得更简单，因此我们应从这两个三角形开始。(要始终设法寻找某些联系。知道两个量相等不见得用得上，除非你通过某种方式把它们联系起来。)由于我们只关心六个点中的四个(A, D, E, O)，所以可以画一张简图来处理问题。(毕竟，为什么一定要应对那些杂乱无用的信息呢?)

我们知道$\angle EAO=\angle DAO=\alpha/2$，而且根据三角形内角和为 180°，可以计算出$\angle AEO=180°-\alpha-\gamma/2=\beta+\gamma/2$；类似地，可以得到$\angle ADO=180°-\alpha-\beta/2=\gamma+\beta/2$。我们还可以标出更多连接点 A, D, E, O 的角，最终得到了如图 4 的简图(为了清晰起见，图被旋转并放大了)。

现在我们可以使用正弦定理了。为了得到一个关于$|OD|$和$|OE|$的有用表达式(这正是我们最初想用正弦定理的原因)，利用正弦定理，我们有

$$\frac{|OD|}{\sin(\alpha/2)}=\frac{|OA|}{\sin(\gamma+\beta/2)}=\frac{|DA|}{\sin(\alpha/2+\beta/2)}$$

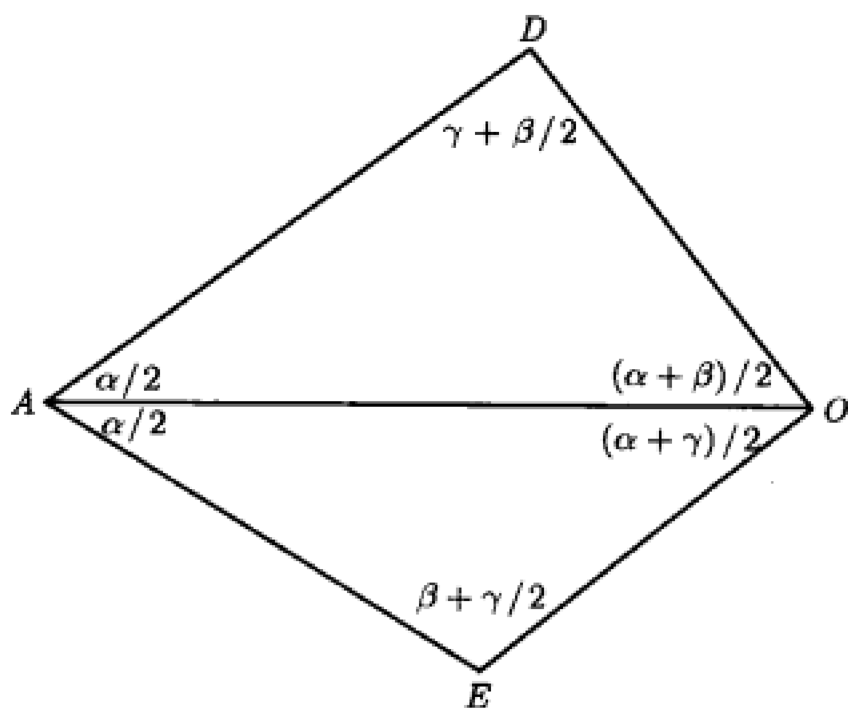


图 4

和

$$\frac{|OE|}{\sin(\alpha/2)}=\frac{|OA|}{\sin(\beta+\gamma/2)}=\frac{|EA|}{\sin(\alpha/2+\gamma/2)}。$$

一旦得到这样两个等式，给定的信息($|OD|=|OE|$)也许就用得上了。边长$|OA|$同时出现在以上两个等式中，所以我们或许应该将$|OD|$和$|OE|$用$|OA|$表达出来，得到

$$|OD|=|OA|\frac{\sin(\alpha/2)}{\sin(\gamma+\beta/2)},$$

$$|OE|=|OA|\frac{\sin(\alpha/2)}{\sin(\beta+\gamma/2)}。$$

因此$|OD|=|OE|$ 当且仅当 $\sin(\gamma+\beta/2)=\sin(\beta+\gamma/2)$。(实际上，可能会出现某些很荒唐的情形，例如 $\sin(\alpha/2)=0$。不难发现，这只发生在极端退化的情形中，而且这些反常的情形可以很容易地单独处理。但是不要忘记这些特殊的情形。)

这里我们已经将有关边的等式转化成为有关角的等式了。更重要的是，这些角与我们的目标(它与角α, β, γ有关)紧密相关。因此，我们确信自己正在正确的方向上前进。我



们的问题现在已经完全转化成了代数问题。

总之，我们得到两个正弦值是相等的，即 $\sin(\gamma+\beta/2)=\sin(\beta+\gamma/2)$。这可以有两种可能性：

$$\gamma+\beta/2=\beta+\gamma/2$$

或

$$\gamma+\beta/2=180^\circ-(\beta+\gamma/2)。$$

我们离目标越来越近了，因为等式中不再含有正弦，而且我们第一次得到了一个含有“或者”一词的命题。不难看出第一种情形可以推出 $\beta=\gamma$，而第二种情形可以得到 $\beta+\gamma=120^\circ$，因而有$\alpha=60^\circ$。不经意之间就已经实现了我们的目标，这可真够奇妙的。

然而，这是真的。有时我们可以很快地把给定的信息转化成一个与目标相关的等式(在这个问题中就是与角α, β和γ相关的等式)，然后运用一些简单的代数技巧将它转化成我们想要的结论。这称为**直接方法**或**前向法**。当目标是一个简单的关系式，只涉及简单计算时，这种方法很有效。这是因为通过逐步简化和转化信息，使之越来越接近于目标，我们就有解决问题的思路了。当目标不是很明确时，我们可能需要先对目标进行转换，然后再确定进行哪种尝试。这正如下一个问题所展示的。

> ***问题 4.3** (Australian Mathematics Competition, 1987, 第 13 页)　设 $ABFE$ 是一个矩形，点 D 是对角线 AF 与 BE 的交点。过点 E 的一条直线交 AB 的延长线于点 G，且交 FB 的延长线于点 C，使得 $|DC|=|DG|$。证明：
>
> $$\frac{|AB|}{|FC|}=\frac{|FC|}{|GA|}=\frac{|GA|}{|AE|}。$$

对于几何问题，可以使用前向法(先系统地找出已知条件，如边和角)或后向法(把最终结果转化为等价但处理起来更简单的结论)。画一张简单的示意图并猜测结论有时会很有帮助，但这个问题的示意图十分难画。你怎么保证$|DC|=|DG|$呢？反复试验(同时兼顾结论$|AB|/|FC|=|FC|/|GA|=|GA|/|AE|$)后，最终画出一张比较合理的示意图(图 5)。

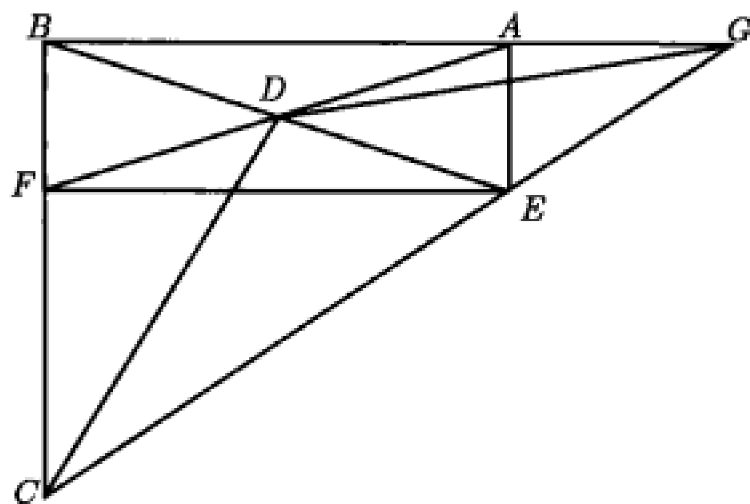


图 5

让我们首先尝试前向法。大刀阔斧型的解析几何法是一种冗长而繁琐的方法，往往会带来无法预测的复杂性并引起巨大的差错，因此，我们把它作为最后的选择(尽管点 A 处的直角看起来是一个非常适合放置原点和坐标轴的位置)。向量方法对于$|DC|=|DG|$之类的关系式也不是很合适(但是向量方法通常比坐标法要简洁)。那么如何来计算线段长度和角度呢？我们知道矩形有四个直角，也知道$|DC|=|DG|$，所以$\triangle DCG$是等腰三角形。但这对我们帮助不大。从点 D 向 CG 作垂线或其他类似的辅助线也与事无补。(接下来我们会看到，某些辅助线的确有用，但在前向法中并不是显而易见的。)

我们再尝试后向法。要证明三个比彼此相等，这就提示我们利用相似三角形。我们能否用某些线段，例如 AB 和 FC，来构造三角形呢？这恐怕不行，但是我们可以用 FE 和 FC



来构造一个三角形，而且 FE 和 AB 是等长的。一旦我们确定了一个三角形，其他两个与之相似的三角形就应该不太难找了。观察图中的$\triangle FCE$我们就可以看出(而且容易证明)它与$\triangle BCG$和$\triangle AEG$是相似的，所以

$$\frac{|EF|}{|FC|}=\frac{|GB|}{|BC|}=\frac{|GA|}{|AE|}。$$

为了与我们要证的结论接近，又可以将上式改写为

$$\frac{|AB|}{|FC|}=\frac{|GB|}{|BC|}=\frac{|GA|}{|AE|}。\tag{15}$$

这样我们已经证明了三个比值中的两个，即$|AB|/|FC|$和$|GA|/|AE|$彼此相等。而我们想要的第三个比值$|FC|/|GA|$很难与某个三角形相联系。但是注意式(15)中间的比值，我们隐约感到这对边与 FC 和 GA 有关。事实上，FC 是 BC 上的线段，而 GA 是 BG 上的线段。这就暗示我们也许证明

$$\frac{|FC|}{|GA|}=\frac{|GB|}{|BC|}$$

比证明

$$\frac{|AB|}{|FC|}=\frac{|FC|}{|GA|}\quad 或\quad \frac{|FC|}{|GA|}=\frac{|GA|}{|AE|}$$

来得容易。此外，前者的表述更对称且只涉及一个等式。

即使有了这个“可能更简单”的式子，看来仍然没有可以利用的相似三角形。此时我们需要对问题做进一步处理。一种显然的想法是尝试重新配置这些比。通过交叉相乘可以得到

$$|FC|\times|BC|=|AG|\times|BG|,$$

或者交换比例项可以得到

$$\frac{|FC|}{|BG|}=\frac{|GA|}{|BC|}。$$

这样做似乎进展不大，但是相乘项$|FC|\times|BC|$和$|AG|\times|BG|$看起来有些眼熟。事实上，我们可能会想起以下定理(它通常出现在高中数学教材中，但却很少被用到)：

定理 4.2 如果点 P 是以点 O 为圆心，r 为半径的圆外一点，从点 P 出发的一条射线交圆于 Q, R 两点，则

$$|PQ|\times|PR|=|PT|^2=|PO|^2-r^2,$$

其中 T 为过点 P 的切线在圆上的切点。

证明 我们观察到$\triangle PQT$ 和$\triangle PTR$ 是相似三角形，因此$|PQ|/|PT|=|PT|/|PR|$(图 6)。而且，由毕达哥拉斯(Pythagoras)定理● 知$|PO|^2=|PT|^2+r^2$，从而定理成立。 □

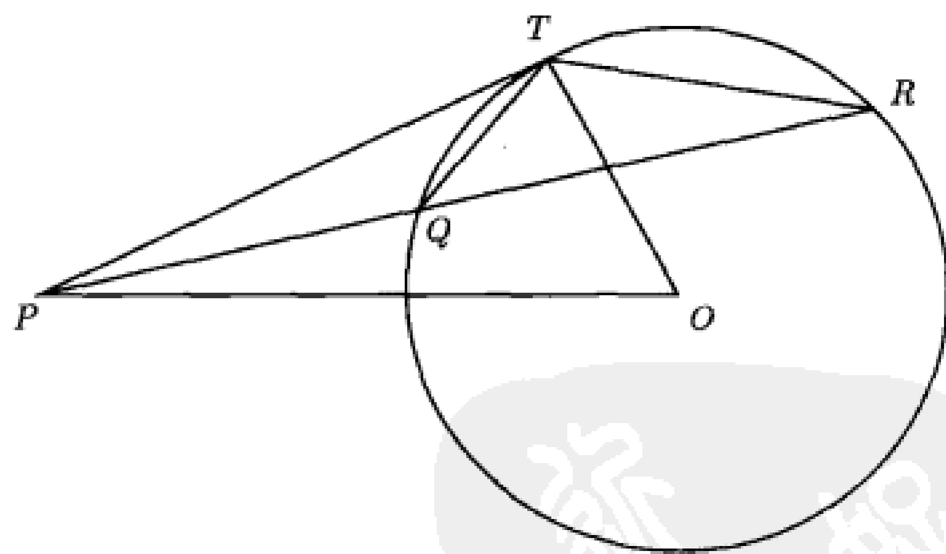


图 6

为了利用定理 4.2，我们首先需要构造一个圆。我们想要计算$|FC|\times|BC|$和$|AG|\times|BG|$，所以这个圆必须包含点 F, B 和 A。现在恰好有一个圆过点 F，B 和 A 并以 D 为圆心。(参考定理 4.1!)所以由定理 4.2 我们有

● 译者注：即勾股定理。



$$|FC| \times |BC| = |DC|^2 - r^2$$

和

$$|AG| \times |BG| = |DG|^2 - r^2,$$

其中 r 是圆的半径。由于 $|DC| = |DG|$ 是给定的条件，所以我们的结论得证。

这是求解纯几何问题的特征：可以利用的信息看来少得可怜，而想要证明的结论却极不易发现，所以通常需要用一种特别的方法来处理。作图或其他辅助手段可能使问题变得清晰，从而触发记忆中某些有用的信息。例如，在某个几何问题中你需要证明 $\angle ABC = \angle ADC$，可代以证明与之等价的结论“四边形” $ABDC$ 有一个外接圆(如果点 C, D 在 AB 同侧)；或者如果你需要证明 $|AB|>|AC|$，可以等价地证明 $\angle ACB>\angle ABC$ (假定点 A, B, C 不共线)；又或者如果是一个关于不同三角形的面积问题，可以利用“等底且等高的三角形面积相等”或“如果三角形的底边减半，则面积也减半”之类的结论。这并不意味着你应把所有可能想到的都在图中构造出来，并罗列一堆已知的结果(除非你实在无计可施)，事实上只要一些有根据的猜测和大体的思路就足够了。有时还可以试用一个特殊或极端的例子来探索问题的解法(例如，在以上的问题中，我们可以考虑四边形 $ABEF$ 是一个正方形，或四边形 $ABEF$ 是退化的，抑或 $|DC|=|DG|=0$ 的情形)。同时要时刻牢记给定的信息(即 $|DC|=|DG|$ 和四边形 $ABEF$ 是一个矩形)和要证的结论 ($|FC|\times|BC|=|AG|\times|BG|$ 或其他表达式)，并设法使你的方法向一些不寻常的条件或目标靠近。(在这个问题中，条件 $|DC|=|DG|$ 看起来有些不寻常。)总之，为了推导出全部结论，我们应该假定需要运用所有的给定条件，所以每个条件都应以某种形式派上用场。

这里的关键之处是要想到欧几里得几何中的某个特殊结论(在这个问题中就是定理 4.2)。在观察问题的各个方面并“抓住”问题的本质后，借助足够多的处理几何问题的经验，这些有用的结论就会在脑海中浮现(常常也会只在所有其他方法都行不通时，这些结论才会在脑海中浮现)。如果没有这样的灵感，我们就应坚持使用解析几何法或者准解析几何法(例如，从点 D 分别向 AB 和 AC 作垂线，并利用毕达哥拉斯定理来表示$|DC|$和$|DG|$。这本质上就是无坐标轴的解析几何法)。

问题 4.4 给定三条平行线，(用直尺和圆规)作一个等边三角形，使得每条平行线各包含三角形的一个顶点。

乍一看，这个问题简单且直接(好题目通常都这样)。但是当我们尝试画一张示意图时(试一试，但要先画出平行线)，就会发现要确定一个三角形并使之满足所有条件是多么需要技巧，因为题目中的要求实在是太苛刻了。在尝试画圆、60°角和类似的图形后，我们意识到这里画图需要一些特殊的技巧。然而，我们还是应该尽力去画一张尽可能好的示意图(也许可以先画出等边三角形，再擦掉它)并标出所有的点和线(图 7)。

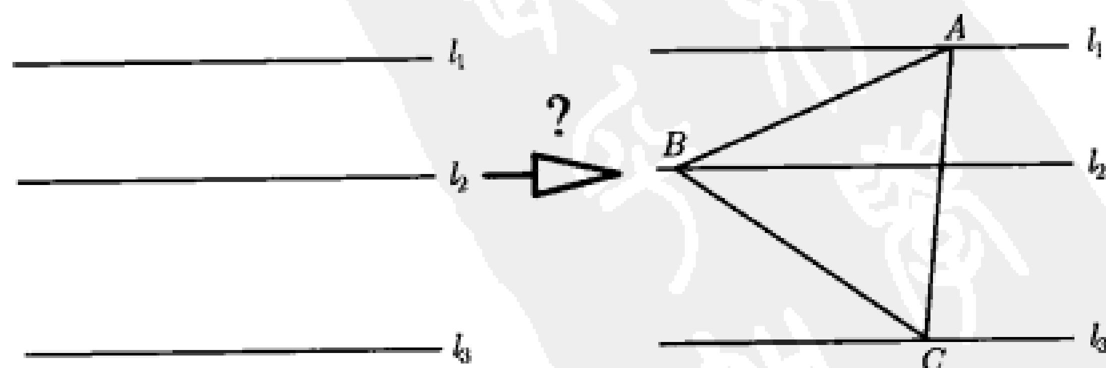


图 7



比较明显的选择是利用解析几何。这种方法可能行得通，但会很麻烦，你将需要用距离公式来定出点的位置，因此这不是最佳(或最具有几何思想的)解法。通常，我们把它留作最后的选择。

解决作图问题的标准方法是取未知量之一(点、线、三角形，或者其他某些量)并确定其轨迹或者它的其他容易作图的性质。但是在这样做之前，让我们先仔细观察这张示意图并做些我们力所能及的尝试。不难发现，一个等边三角形，如果存在的话，可以在平行线上滑动并仍然满足所有的要求。因此，如果$\triangle ABC$是这样一个三角形，那么只要点 A 在直线 l_1 上，它的位置就是任意的。当然，点 B 和 C 将依赖于点 A 的位置。所以本质上可以把点 A 放置在任意我们想要放的位置上，而不必担心它失去一般性，然后只需关注点 B 和 C 就可以了。这些讨论表明直线 l_1 已经变得无关紧要。它原本只用来约束点 A，但一旦我们把点 A 固定在 l_1 的任意一点处后，就不再需要 l_1 了。

现在，随着点 A 被确定，$\triangle ABC$ 受到了更多的限制。也许这种限制会使得点 B 和 C 只有有限种位置的选择，但目前为止我们还不是很清楚。

等边三角形 ABC 现在只有两个自由度：方向性和尺寸。但是它还有两个限制：点 B 一定在直线 l_2 上以及点 C 一定在直线 l_3 上。从理论上讲这些应足以固定三角形了，但对于像三角形这样复杂的几何图形(相对于点和线)来说，很难看出接下来该怎么办。然而我们可以做的是把一个未知量转化为另一个较易处理的未知量。目前的未知量就是等边三角形。什么是较简单的未知量呢？最简单的几何图形是点。因此，我们可以先确定一个点，例如点 B，而不是去确定整个三角

形。由于点 B 被限制在 l_2 上，所以它只有一个自由度。点 B 由什么条件来确定呢？此条件就是：以 AB 为边的等边三角形的第三个顶点(即点 C)一定在 l_3 上。这个条件还是很复杂，因为它仍然涉及等边三角形。是否有一种更简单的方法将点 C 用点 A 和 B 来表示呢？答案是肯定的：点 C 是由点 B 绕着点 A 旋转 60°(沿着顺时针或逆时针方向)得到的。因此问题简化为：

> 给定点 A 和不经过点 A 的两条平行线 l_2, l_3，找出 l_2 上的一点 B，使得点 B 在绕着点 A 旋转 60°后落在 l_3 上 (图 8)。

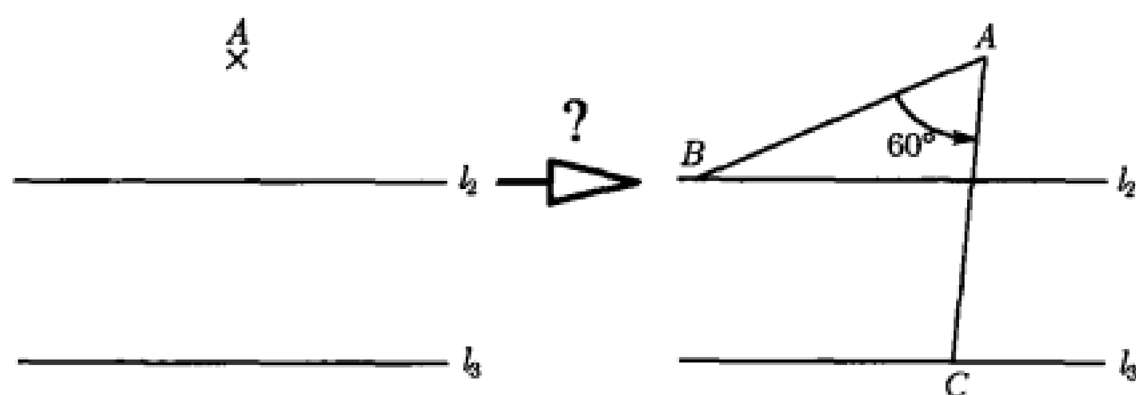


图 8

我们现在只有一个未知量——点 B，所以自由度更少了，问题也应更简单了。我们要点 B 满足两条性质：

(a) 点 B 在 l_2 上；

(b) 点 B 绕着点 A 旋转 60°后落在 l_3 上。

条件(b)是一种不太好用的形式，除非你把它转化为：

(b') 点 B 落在由 l_3 绕着点 A 反向旋转 60°后的直线上。也就是说，点 B 落在 l_3' 上，这里 l_3' 是由 l_3 绕着点 A 反向旋转 60°(顺时针或逆时针)得到的。

于是我们的性质变为：



(a) 点 B 在 l_2 上；

(b') 点 B 在 $l_{3'}$ 上。

或者换句话说，点 B 是 l_2 和 $l_{3'}$ 的交点。这就对了！我们已经明确地作出了点 B，所以，这个三角形就很容易得到了。

为了保持完整性，这里给出整个作图过程：

> 选取 l_1 上任意一点 A。将 l_3 绕着点 A 旋转 60°(顺时针或逆时针；对于每个给定的点 A，点 B 都有两个解)，并设旋转后的直线与 l_2 的交点为点 B。再将点 B 反向旋转 60° 得到点 C。

请注意这一作图法也适用于不是平行线的情形，只要它们之间的夹角不是 60° 就可以。因此“平行性”事实上只是用来分散注意力而已！

与代数问题一样，作图问题的解题思路也是“求解”一个未知量，在这个问题中就是点 B。我们不断地对已知信息进行转化，直到它具有“点 B 是……”的形式。为了给出一个类似的代数问题，我们来看以下的例子：假定我们要从以下给定的信息中解出 b 和 c：

- $b+1$ 是偶数；
- $bc=48$；
- c 是 2 的幂。

如果我们从以上三个给定条件中解出 b 并消去 c，则得到

- b 等于偶数减 1(即 b 是奇数)；
- b 等于 48 除以 2 的某次幂(即 b=48, 24, 12, 6, 3, 1.5, …)。

然后通过比较奇数集和 48 除以 2 的幂所得到的所有数的集合，我们发现 $b=3$。用逐个消元法求解多变量问题通常比较容易，此方法在几何作图问题中同样有效。

习题 4.1 设 k 和 l 是两个圆，相交于点 P 和 Q。试作一条过点 P 但不过点 Q 的直线 m，使其满足：如果 m 交 k 于点 B 和 P，交 l 于点 C 和 P，那么 $|PB|=|PC|$（图 9）。（提示：找出点 B。）

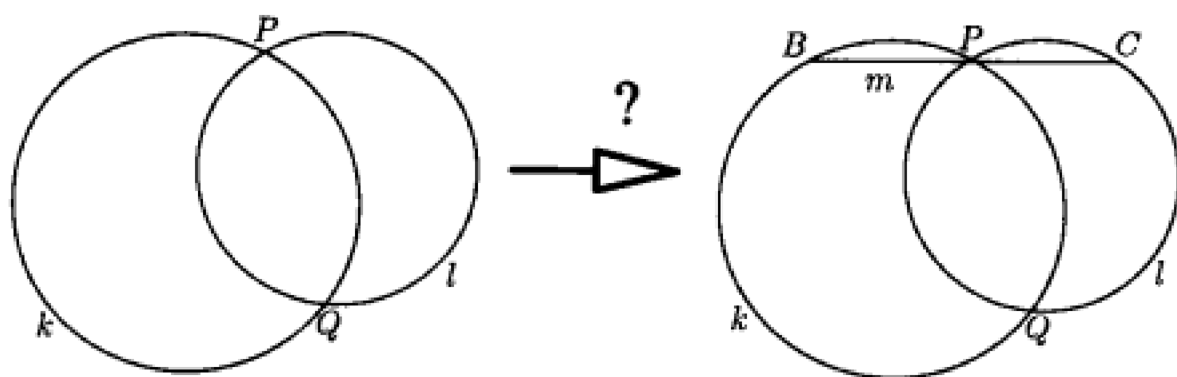


图 9

习题 4.2 给定一个圆以及圆内两点 A 和 B。如果可能的话，作圆的一个内接直角三角形，使其一条直角边包含点 A，另一条直角边包含点 B（图 10）。（提示：找出直角所在的顶点。）

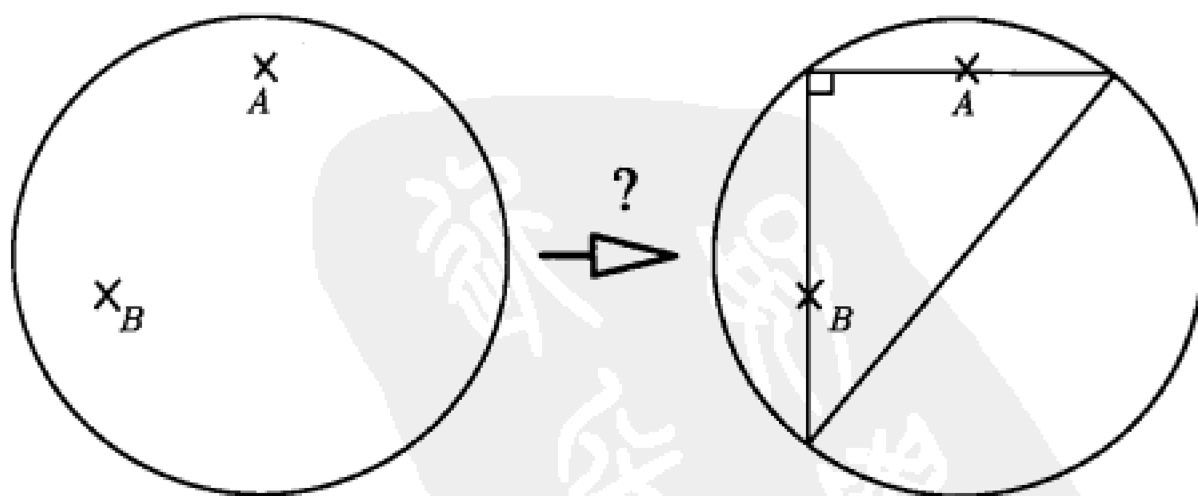


图 10

*__习题 4.3__ 给定四个点 A, B, C 和 D。如果可能，作一个正方形，使其每条边分别包含这四个点中的一个



点(图 11)。(提示：遗憾的是，作这个正方形非常困难，即使只要找出这个正方形的一个顶点(如同我们在前面的问题中所做的那样)也简单不了多少，因为只知道这个顶点被限制在一个固定的圆上，但也仅此而已。一种解决问题的方法是确定这个正方形的一条对角线。一条对角线依赖于若干要素：方向、位置和端点。但是对角线可以唯一确定正方形，而这是单独一个顶点不容易做到的。如果你实在无计可施，就去画出一张漂亮的大示意图：先画出正方形，再画出 A, B, C, D 四个点，然后分别画出以 AB, BC, CD 和 DA 为直径的圆，再画出两条对角线。充分利用这些圆的有利条件来计算角度，找出相似三角形，等等。一个真正重要的提示是，观察对角线与圆的交点。另一种方法是确定一条特定的边，这需要利用旋转、反射和平移来将一条边转变到与另外一条边基本重合。简而言之，这是与上述类似的一种解法。)

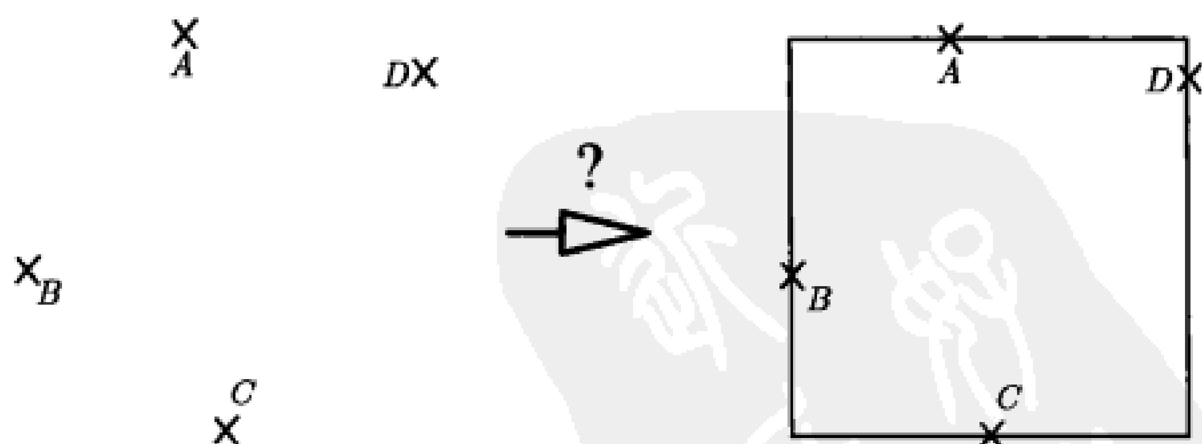


图 11

问题 4.5 (Taylor, 1989，第 10 页，问题 4) 一个正方形被分成五个矩形，如图 12 所示。四个外围矩形 R_1, R_2, R_3, R_4 面积相等。证明内部的矩形 R_0 是正方形。

这又是一个"结论不同寻常"的问题。乍一看，外围的四个矩形面积相等这一事实似乎并不一定使得内部的矩形是等边的。首先你可能感到给定的条件有太多的自由度，毕竟面积固定的矩形可以是长而窄的，也可以是短而宽的。我们为什么不能调整外围矩形的形状使得内部的矩形变形呢？一个简单的尝试就能说明为什么这样行不通：每个矩形都被其相邻的矩形所约束。例如，在图 12 中，矩形 R_1 被矩形 R_2 和 R_4"限制在适当的位置上"，改变矩形 R_1 将导致矩形 R_2 和 R_4 的变化，这都将使得 R_3 改变。但是矩形 R_3 不能同时满足矩形 R_2 和 R_4 的要求，除非它们对 R_3 的要求相同。在图 13 中，矩形 R_3 可以与矩形 R_2 相配，或者与矩形 R_4 相配，但不能与它们同时都相配(请记住 R_3 还需要和 R_2, R_4 具有相等的面积)。我们开始意识到该如何"处理"这问题了：因为外围矩形需要相等的面积，以及"齐平拼接"的困难，所以能够这样处理的唯一可能的办法就是内部矩形为正方形。我们不可能避开这个对称的"卍字形"[②]结构，图 13 就给出了一个为什么可能出错的例子。

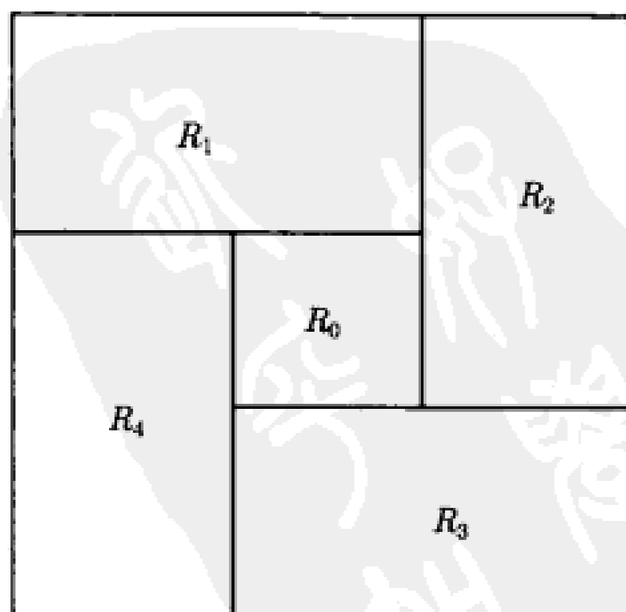


图 12



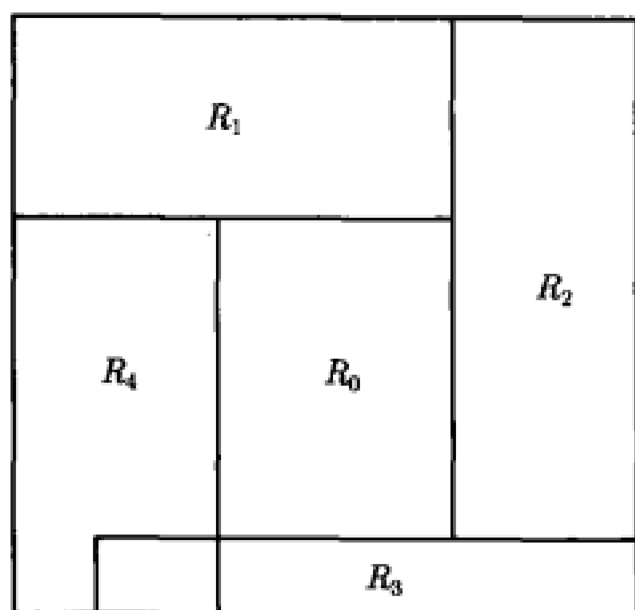


图 13

为了方便进一步的推导，我们需要引进一些符号，更具体地讲，需要用较少的变量来表示所有几何对象的尺寸和面积。从我们对这一结构作“变动”的讨论可以看出，一个矩形，例如矩形 R_1，将决定所有其他矩形的位置：R_1 将迫使 R_2 和 R_4 处于确定的位置，从而也将确定 R_3，如果可能的话。于是我们就有了一种代数表达方法：假设矩形 R_1 的尺寸是 $a\times b$，大正方形的边长为 1，我们来确定所有其他矩形，特别是 R_0 的尺寸。这种方法十分有效，我们最终将会得到关于 R_3 的两个方程(如果处理的方式不同，也可能是得到关于 R_1, R_2 或 R_4 的方程)，从而可以得到 a 和 b 的一个关系式(并非 R_1 的任何边长都行得通，事实上我们需要证明只有那些使得中间的矩形为正方形的 R_1 才是被允许的)。图 14 对以上讨论做了小结。

为了使 R_3 有所要的面积，需满足

$$\left(1-\frac{ab}{1-a}\right)\times\left(1-\frac{ab}{1-b}\right)=ab\text{。}$$

利用这个方程，可以解出 a, b，从而确定 R_0 是正方形。这种方法是可行的，但代数计算有点儿繁琐。因此让我们尝试一种更简单、更直观且更少依赖于坐标的方法(实际上，这种方

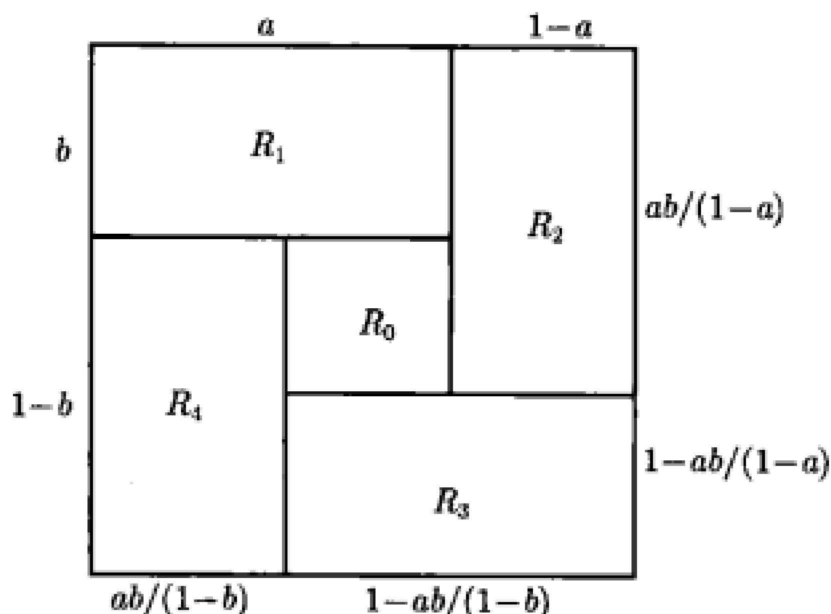


图 14

法的主要思想都是利用坐标)。

我们想要证明只有当 R_0 是正方形时所有的条件才能得到满足，但证明这一点有些难度。我们已经说明了可以把所有的量用矩形 R_1 的边长来表示。从这种意义上讲，矩形 R_1 可以被称为“主要图形”：所有其他的构形都必须依赖于它。一旦我们有了这个参考点，就可以把注意力集中到一个矩形上。因为 R_0 不像其他矩形那样容易成为“主要图形”，所以我们并不想去证明有关矩形 R_0 的性质。但我们可以证明有关矩形 R_1 的性质，这要来得更容易。

图 14 似乎暗示 $a+b$ 应该等于 1。的确，如果 $a+b=1$，那么 R_2 的水平边长一定为 $1-a=b$，由面积相等得到竖直边长为 a，于是矩形 R_3 的竖直边长一定为 $1-a=b$，依此类推。这与前面提到的“卍字形”结构非常吻合，而且可以导出 R_0 是边长为 $a-b$ 的正方形。这样我们就提出了一个中间目标：证明 $a+b=1$。因为所有量都可以用 a 和 b 来表示，所以启发我们希望这个目标比较容易实现。然而，要用矩形 R_0 的边长来表示所有量就不容易了。



概括地讲，我们已经证明了以下推导链中的第二个蕴涵关系：

$$\boxed{R_1, R_2, \cdots, R_4\text{面积相等}} \to \boxed{a+b=1} \to \boxed{R_0\text{是正方形}}$$

现在只需证明第一个蕴涵关系就可以了。

由解析几何方法我们知道，尽管很容易把给定的条件转化为等式，但从这些等式推出所要的结论并不容易。虽然面积相等看起来是非常漂亮、简单且易于处理的条件，但因为我们只有一系列相等的乘积，且它们中的项还和另外的关系式有关，所以实际上它们更多地是成为问题的障碍。但是我们可以反过来讨论，即设法证明：

$$\boxed{a+b\neq 1} \to \boxed{R_1, R_2, \cdots, R_4\text{面积不相等}}$$

或者用反证法证明：

$$\boxed{a+b\neq 1}\ \text{和}\ \boxed{R_1, R_2, \cdots, R_4\text{面积相等}} \to \boxed{\text{矛盾}}$$

请注意，在用反证法时，我们可以从更多的信息入手，但最终的结果是没有限制且不确定的。这种策略非常适合前面所采用的定性方法，因为我们不可能去变动对称的结构，否则所有的矩形会失去平衡。因此让我们将更多的精力集中到反证法上。

假设 $a+b$ 过大，即 $a+b>1$，但这四个矩形以某种方式达到面积相等的要求。这时我们要找到一个矛盾。如果我们有一个较大的矩形 R_1，那会怎么样呢？它将影响相邻的矩形，比如说影响 R_2，使它变“窄”。实际上，R_2 的水平边长为 $1-a$，这个值小于 R_1 的竖直边长 b。所以 R_2 比 R_1 变“窄”了。因为有矩形面积相等这一限制，所以 R_2 的竖直边一定比 R_1 的水平边长。因此，R_2 的两边之和也大于 1，且 R_2 比 R_1 变“宽”了[20]。再看 R_3：根据相同的逻辑推理，R_3 的竖直边要比 R_2

的水平边短，R_3 的水平边一定比 R_2 的竖直边长。因此，R_3 的两边之和也大于 1，且 R_3 也比 R_2 变“宽”了。再利用同样的推理可知，R_4 的水平边(或竖直边)要比 R_3 的竖直边(或水平边)短(或长)。因此，R_4 的两边之和也大于 1，且 R_4 一定比 R_3 更“宽”了。最后可推出：R_1 的水平边(或竖直边)要比 R_4 的竖直边(或水平边)长(或短)。而这意味着 R_1 的水平边(或竖直边)要比它自己的水平边(或竖直边)长(或短)。这是荒谬的，我们找到了矛盾。当 $a+b<1$ 时也会产生类似的情形：我们最终可以推出 R_1 的水平边(或竖直边)要比它自己的水平边(或竖直边)短(或长)。这也是一个谬论。

这个问题是说明“一张图比 1000 个关系式更有价值”的极好范例。而且，请记住有时使用不等式比使用等式更简单、更有效。

习题 4.4 找出满足

$$x^p+y^q=y^r+z^p=z^q+x^r$$

的所有正实数 x, y, z 和正整数 p, q, r。(提示：这个问题不涉及几何知识，但它的解法仍与问题 4.5 类似。)

问题 4.6 (AMOC Correspondence Problem，1986—1987，第 1 集，问题 1) 设 $ABCD$ 是一个正方形，k 是以点 B 为圆心且过点 A 的圆，l 是正方形内以 AB 为直径的半圆。再设点 E 是 l 上的一点，BE 的延长线交圆 k 于点 F。证明：$\angle DAF=\angle EAF$。

和通常一样，我们先画一张示意图(图 15)。我们需要证明两个角相等。鉴于题目中缺乏边长等信息，看起来我们完全可以通过角来处理问题。毕竟，圆总是与角紧密联系的。但是$\angle DAF$和$\angle EAF$这两个特定的角看起来关系不大，因此



我们需要把这两个难处理的角用“关系更密切的”角表示出来，使得可以在这两个角之间建立联系。

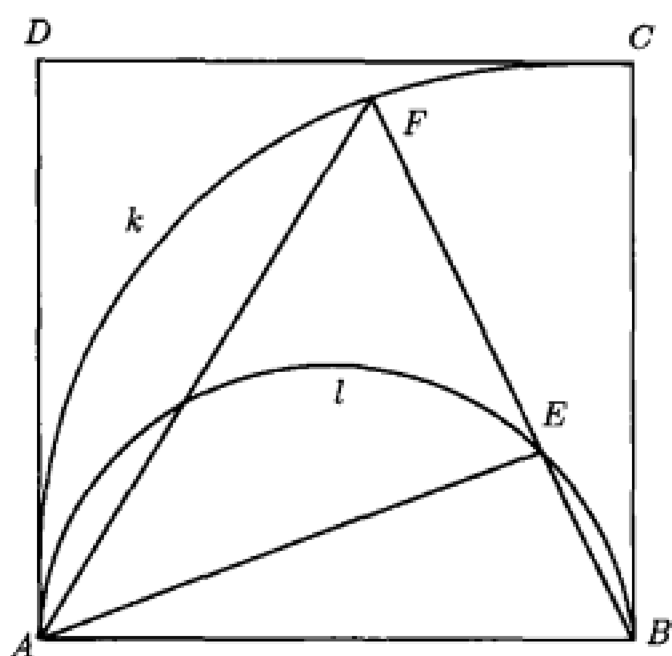


图 15

让我们从$\angle DAF$入手。$\angle DAF$不与任一三角形关联，但它与圆 k 有关。这里可以运用一个古老的小定理(欧几里得 III, 32)[⑬]，即一条弦所对应的圆周角与其弦切角相等。因此，我们可以说$\angle DAF=\angle APF$，这里点 P 是圆 k 上的任意一点，它在包含点 C 的弧 AF 上。例如，我们可以说$\angle DAF=\angle ACF$，尽管$\angle ACF$和$\angle DAF$一样不太顺眼。但是它是一个圆周角，这意味着它是同一条弦所对应的圆心角的一半，也就是说$\angle ACF=(1/2)\angle ABF$。与某些三角形和圆关联的$\angle ABF$看起来是一个更“主流的”角，所以$\angle DAF=(1/2)\angle ABF$是一个令人满意的结果。

现在我们开始研究$\angle EAF$。可是这个角甚至比$\angle DAF$更难处理，因为它与任何其他角都没有直接关系。然而它与其他较好用的角，例如$\angle DAB$, $\angle EAB$ 等，有一个公共顶点，所以我们可以把$\angle EAF$用与其关系密切的角来表示，例如

$$\angle EAF = \angle BAF - \angle BAE,$$

或者也可以是

$$\angle EAF = \angle DAB - \angle DAF - \angle BAE.$$

第一个等式带给我们一个用得上的角$\angle BAE$和一个不太好利用的角$\angle BAF$。然而，第二个等式有更多的优越性：由于$\angle DAB=90°$，而且我们已经解出了$\angle DAF$，所以得到

$$\angle EAF = 90° - \frac{1}{2}\angle ABF - \angle BAE。$$

而$\angle BAE$和$\angle ABF$都在$\triangle ABE$中，既然我们已经把$\angle DAF$和$\angle EAF$都用$\triangle ABE$中的角表示出来，很明显地就该把注意力集中到这个三角形上。

$\triangle ABE$内接于一个半圆。这提醒我们使用泰勒斯定理(定理 4.1)。此定理告诉我们$\angle BEA=90°$。因为三角形的内角和为 180°，所以这就把$\angle ABF$和$\angle BAE$联系起来了。准确地说，我们有$\angle ABF+\angle BAE+\angle BEA=180°$，所以$\angle BAE = 90°-\angle ABF$。现在我们把这个式子代入$\angle EAF$的表达式，得到

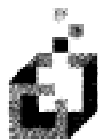
$$\begin{aligned}\angle EAF &= 90° - \frac{1}{2}\angle ABF - \angle BAE \\ &= 90° - \frac{1}{2}\angle ABF - (90° - \angle ABF) = \frac{1}{2}\angle ABF。\end{aligned}$$

这与前面得到的$\angle DAF$的表达式是相同的。因此我们证明了$\angle EAF=\angle DAF$。当然，在写出证明时我们需要对以上过程加以整理。我们可能希望得到如下的一串连等式：

$$\begin{aligned}\angle DAF &= \cdots \\ &= \cdots \\ &= \cdots \\ &= \angle EAF,\end{aligned}$$



但是在寻求解法时，不必写得如此正式。如果你知道自己要找的是什么，那么计算出$\angle DAF$和$\angle EAF$，并希望它们与某个中间量相等就是一个不错的想法。只要不断努力简化问题并建立一些联系，解决问题的机会很快就会出现。(当然，这里我们假定问题有解，而大多数问题并不会用“无解”来戏弄你。)



第五章 解析几何

几何的思维并不仅仅限于几何学，它可以脱离几何学而用于其他知识领域。在其他条件等同的情况下，只要假以几何思维之手，无论是有关伦理道德、政治、评论，甚或口才方面的事都会被做得更加优雅完美。

——Bernard le Bovier de Fontenelle

本章中的问题涉及几何的概念和研究对象，但这些问题的解答需要用到其他数学分支如代数、不等式、归纳法等的思想。为了应用向量算术的定理，有时行之有效的一种技巧是利用向量来重新表述几何问题。这里给出一个这样的例子。

问题 5.1 (Australian Mathematics Competition, 1987, 第 14 页) 一个正 n 边形内接于一个半径为 1 的圆。设 L 是由连接多边形顶点的所有线段所可能有的不同长度组成的集合。问：L 中所有元素的平方和是多少？

首先，让我们为“L 中所有元素的平方和”起一个更短的名字，例如“X”；之后我们的任务就是去计算 X。这是一个所谓“可行的”问题：它既不是一个“证明……”型的问题，也不是“是否存在……”型的问题，而是要去算出一个数，比如说通过直接应用三角学知识和毕达哥拉斯定理来得到。例如，当 $n=4$ 时，我们就得到单位圆(半径为 1)的一个



内接正方形，其中顶点间连线的可能的不同长度为：边长 $2^{1/2}$，对角线长 2，所以 $X=(2^{1/2})^2+2^2=6$。类似地，当 $n=3$ 时，我们得到的唯一的顶点间连线长度是边长，即 $3^{1/2}$，所以在这种情形中 $X=(3^{1/2})^2=3$。当 $n=5$ 时，推算就不那么简单了，除非你知道若干正弦和余弦值。所以让我们跳过这种情形而来尝试 $n=6$。这时边长为 1，较短的对角线长是 $3^{1/2}$，较长的对角线(圆的直径)长是 2，所以在这种情形中 $X=1^2+(3^{1/2})^2+2^2=8$。最后我们考虑退化的情形：$n=2$。在这种情形中“多边形”只是一条直径，所以 $X=2^2=4$。于是，我们就计算出了一些特殊情形的 X 值，见表 1，其中我们用“？”标出 $n=2$ 的情形，因为说“一个两条边的多边形”有点儿牵强。

表 1

N	X
2?	4?
3	3
4	6
6	8

这个小表格并不能对一般情形的解答提供太多线索。下面我们首先要做的是画一张示意图，并在示意图上把顶点标出，这可能使问题看得清楚些。对于 n 的某个固定值(例如 $n=5$ 或 $n=6$)，顶点可以标为 A, B, C 等，但对于一般的情形，把顶点标为 $A_1, A_2, A_3, \cdots, A_n$ 可能更方便，如图 1 所示。

现在我们可以进行初步的观察和猜测：

(a) 对于 n 是奇数或偶数，情形可能是不同的。如果 n 为偶数，我们有长的对角线。事实上，当 n 为偶数时我们有 $n/2$ 条长度不同的顶点连线，而当 n 为奇数时有 $(n-1)/2$ 条长度不同的顶点连线。

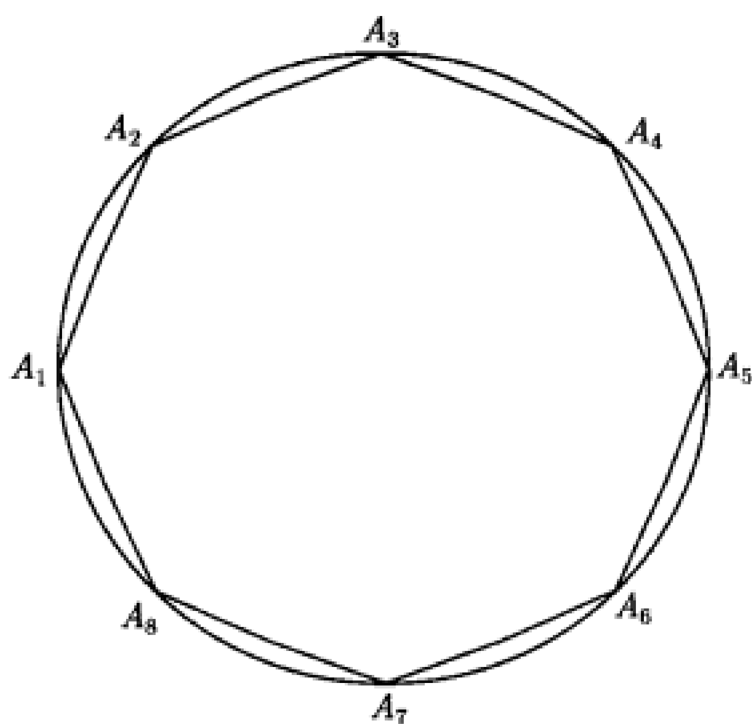
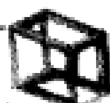


图 1

(b) 问题的答案可能总是一个整数。目前这并不是一个很确定的猜想，因为我们讨论的是一些非常特殊的等边三角形、正方形及六边形等，它们的边长都是平方根型的。然而，这使我们感到问题的一般解也许比较整齐。

(c) 我们要计算的是长度的平方和，而不是长度本身的和。这就使我们马上放弃利用纯几何的求解想法，转而去考虑解析几何。解析几何让我们想到向量或坐标几何，抑或复数(实质上它们都是同样的方法)。对于涉及三角和的问题，坐标几何方法是一种并非快捷但很可靠的方法，而向量几何和复数这两种方法看起来更有利用价值(向量几何法可以利用点乘，而复数方法可以利用复指数)。

(d) 因为我们计算的不是所有对角线长度的平方和，而只是所有不同长度的对角线的长度的平方和，所以试图直接解决问题几乎是不可能的。但是我们可以重新叙述问题，使之具有更容易转化为方程的形式。(方程是可靠的数学工具，虽然不像示意图和解题思路那样富有启发性，但最容易操作。



通常，除了某些组合学和图论中的例外，我们总是把要求的目标表示为某类方程。)不管怎样，如果你只考虑从多边形的某个定点出发的所有对角线，那么这些对角线将包括我们需要的所有长度。

例如，在图 2 中，n 是偶数，我们有四条长度不同的顶点连线。如果你只注意上半圆，那么对角线的每个长度恰好出现一次，即长度$|A_1A_2|$，$|A_1A_3|$，$|A_1A_4|$和$|A_1A_5|$将包括我们想要的所有长度。换句话说，答案可以表示成一个表达式：$|A_1A_2|^2+|A_1A_3|^2+|A_1A_4|^2+|A_1A_5|^2$。在更一般的情形中，我们就要计算$|A_1A_2|^2+|A_1A_3|^2+\cdots+|A_1A_m|^2$，这里 $m=n/2+1$(如果 n 是偶数)或 $m=(n+1)/2$(如果 n 是奇数)。所以我们对这个问题可给出一种更明确的表述：

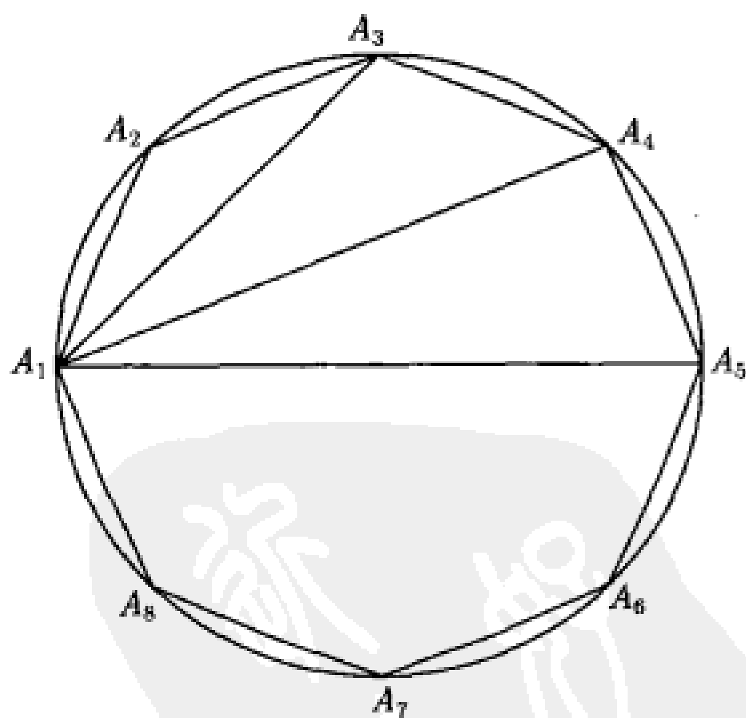


图 2

> 设一个具有 n 个顶点 A_1，A_2，…，A_n 的正多边形内接于半径为 1 的圆中。如果 n 为偶数，令 $m=n/2+1$；如果 n 为奇数，令 $m=(n+1)/2$。计算 $X=|A_1A_2|^2+|A_1A_3|^2+\cdots+|A_1A_m|^2$ 的值。

和式$|A_1A_2|^2+|A_1A_3|^2+\cdots+|A_1A_m|^2$的末项为$|A_1A_m|^2$，而不是更自然的$|A_1A_n|^2$，这使得计算不太方便。但是我们可以作“加倍”处理(像在问题 2.6 中一样)。利用对称性我们有$|A_1A_i|=|A_1A_{n+2-i}|$，从而有

$$X=\frac{1}{2}(|A_1A_2|^2+|A_1A_3|^3+\cdots+|A_1A_m|^2+|A_1A_n|^2$$
$$+|A_1A_{n-1}|^2+\cdots+|A_1A_{n+2-m}|^2)。$$

请注意，当n是偶数时，我们计算了两次对角线项$|A_1A_{n/2+1}|^2=4$。对上式加以整理，同时为了保证对称性，再加上$|A_1A_1|^2$(它等于 0)。所以，当n是奇数时，得到

$$X=\frac{1}{2}(|A_1A_1|^2+|A_1A_2|^2+\cdots+|A_1A_n|^2); \tag{16}$$

而当n是偶数时，得到

$$X=\frac{1}{2}(|A_1A_1|^2+|A_1A_2|^2+\cdots+|A_1A_n|^2)+2 \tag{17}$$

(最后多出来的一项“2”出自对角线项$|A_1A_{n/2+1}|^2=4$与 1/2 相乘的结果)。这样，很自然地，我们引进新的变量

$$Y=|A_1A_1|^2+|A_1A_2|^2+\cdots+|A_1A_n|^2, \tag{18}$$

并试图计算Y，而不是X。这样做的优越性在于：

- 一旦知道Y，等式(16)和(17)立刻给出X；
- Y具有比X更漂亮的形式，因此可能更容易计算；
- 当计算Y时，我们不必区分n是偶数或奇数的两种情形，可以节省一些工作量。

我们再回到前面关于小的$n=3, 4, 6$的表 1，并利用式(16)和(17)计算对应的Y值，得到表 2。从这个表格我们可以推测$Y=2n$。式(16)和(17)意味着，当n是奇数时$X=n$，而当n是偶数时$X=n+2$。这可能就是我们要找的答案，但仍需要去



证明它。

表 2

n	X	Y
2?	4?	4?
3	3	6
4	6	8
6	8	12

现在该用到向量几何了，因为它能够给我们提供一些有用的工具来处理类似式(18)的表达式。因为向量 $\boldsymbol{v}$ 的长度的平方可以简单地表示为它与自身的点乘 $\boldsymbol{v}\cdot\boldsymbol{v}$，所以我们可以把 Y 写为

$$Y=(A_1-A_1)\cdot(A_1-A_1)+(A_1-A_2)\cdot(A_1-A_2)+\cdots+(A_1-A_n)\cdot(A_1-A_n),$$

这里我们把 $A_1, A_2,\cdots, A_n$ 看成是一些向量而不是一些点。坐标原点可以选择在我们需要的任意位置上，但最符合逻辑的选择是以圆的中心作为原点(其次是选择 A_1 作为原点)。把原点放在圆心处，立竿见影的优势是所有的向量 $A_1, A_2,\cdots, A_n$ 都具有长度 1，这样 $A_1\cdot A_1=A_2\cdot A_2=\cdots=A_n\cdot A_n=1$。特别是，我们可以利用向量算术得到

$$(A_1-A_i)\cdot(A_1-A_i)=A_1\cdot A_1-2A_1\cdot A_i+A_i\cdot A_i=2-2(A_1\cdot A_i),$$

所以可以把 Y 展开为

$$Y=[2-2(A_1\cdot A_1)]+[2-2(A_1\cdot A_2)]+\cdots+[2-2(A_1\cdot A_n)]。$$

我们可以合并某些项并化简，得

$$Y=2n-2[A_1\cdot(A_1+A_2+\cdots+A_n)]。$$

我们已经猜测 $Y=2n$，所以如果我们可以证明向量和 $A_1+A_2+\cdots+A_n=\mathbf{0}$，那么就证明了这个猜想。根据对称性，这是显然的(向量以大小相等的力在各个方向上“拉”，所以合成

的净结果一定是 **0**。换一个角度，你可以说正多边形的质心与其中心重合。注意到我们总是要充分利用对称性)。所以 $Y=2n$，从而证明了当 n 为奇数时 $X=n$，而当 n 为偶数时 $X=n+2$。

也许有人对运用对称性轻描淡写地说明 $A_1+A_2+\cdots+A_n=\mathbf{0}$ 不满意。我们可以利用三角学或复数给出一个更具体的证明，但这里还有一个更清晰的对称论证方法，可能使你更满意：记 $\boldsymbol{v}=A_1+A_2+\cdots+A_n$。把整个平面绕着原点旋转 $360°/n$，这使得所有的顶点 $A_1, A_2, \cdots, A_n$ 依次移到下一个顶点处，但并没有改变向量和 $\boldsymbol{v}=A_1+A_2+\cdots+A_n$。换句话说，当我们把 $\boldsymbol{v}$ 绕着原点旋转 $360°/n$ 时，我们得到的仍是 $\boldsymbol{v}$。唯一能使这成立的只可能是 $\boldsymbol{v}=\mathbf{0}$，从而 $A_1+A_2+\cdots+A_n=\mathbf{0}$。这正是我们想要证明的结论。

对以上论证，我们可给以物理解释。实际上，平方和 Y 实质上就是关于 A_1 的转动惯量，这样就可以利用平行轴的施泰纳(Steine)定理(Borchardt，1961，第 370 页)将旋转点移到重心[②]。

习题 5.1 证明：单位立方体在任意平面上的投影的面积等于这个立方体在该平面的垂线上的投影的长度。(提示：本题有一种简洁的向量解法，但它需要对向量叉乘及相关运算有好的处理方法。首先，选择一个恰当的坐标系，并选择一些可以找到的最容易处理的向量。然后有效地利用向量的叉乘、点乘以及大量成对的垂直向量所具有的优势，写出本题要讨论的量并对其进行运算。其中也要利用许多向量 $\boldsymbol{v}$ 具有单位长度这一事实(这使得 $\boldsymbol{v}\cdot\boldsymbol{v}=1$)。最终，一旦找到了解决方法，你就可以重写证明过程，并观察到向量解法是多么的精练、简洁。)



> *问题 5.2　一个矩形被分割成若干个小矩形(图 3)，每个小矩形都至少有一条长度为整数的边。证明：原来的大矩形至少有一条长度为整数的边。

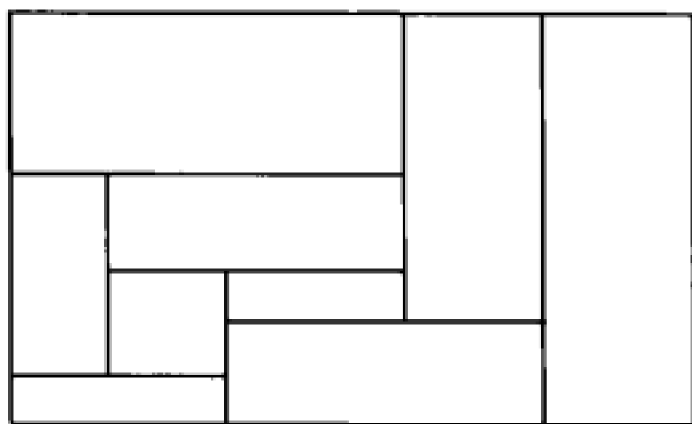


图　3

这是一个看起来很有趣的问题，想必会有一种令人满意的解法。但是结论有点儿奇怪：如果所有小矩形都有一条(或者也许是更多条)长度为整数的边，那么为什么大矩形也该有一条长度为整数的边呢？如果我们的研究对象不是矩形，而只是线段，问题就简单了：长线段是由长度均为整数的短线段组成，所以长线段的长度就是整数之和，当然也是一个整数。一维情形不一定直接对二维情形提供明显的帮助，除了提示我们应该利用整数之和是整数这一事实。于是，我们可以立即引进一个便于叙述的概念：称长度为整数的边为“整数边”。

然而，主要是因为题目中有“分割”一词，所以这个问题也许与拓扑学、组合学，甚至更复杂的数学分支有联系。这样说有点儿太笼统了。为了掌握这个问题，让我们从下面最简单(但并不平凡)的分割着手(图 4)。

假定我们有两个小矩形，并且知道每个都至少有一条整数边。但是那些整数边可能是水平的，也可能是竖直的，我

们无法确定。假设左边的子矩形有一条竖直的整数边，因为这条边的长度与大矩形的竖直边长相等，所以就证明了大矩形有一条整数边。

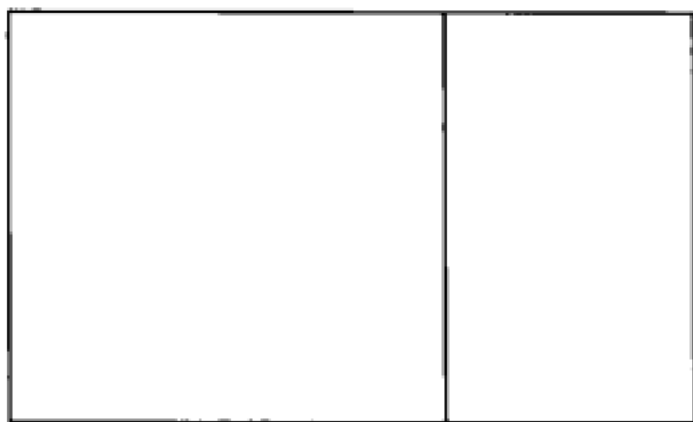


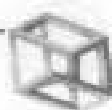
图 4

于是我们可以假设左边的小矩形有一条水平的整数边。由类似的推理我们可以假设右边的小矩形有一条水平的整数边。因为大矩形是两个具有水平整数边的小矩形之和，所以它也有一条整数边。因此我们对于分割为两个小矩形的特殊情形证明了这个问题。但是这个证明是如何实现的呢？(只有当例子能揭示出如何去处理一般问题时，它才是真正有价值的)。反复查看以上的证明，我们观察到下面两个要素：

(a) 因为每个小矩形可能有一条竖直的或水平的整数边，所以我们需要分两种情形考虑。

(b) 可以说，证明大矩形有一条，比如说，竖直整数边的唯一途径是，找到“一串”小矩形，它们都有一条竖直的整数边，并且这些小矩形的竖直边可以通过某种方式“合成”成大矩形的竖直边[22]。这里给出了一个例子，图 5 中被阴影覆盖的小矩形各有一条水平的整数边，从而大矩形也就有一条水平的整数边[23]。

所以，根据这些模糊的想法，我们可以提出以下不很明



确的策略：

> 找到一串具有水平整数边的矩形或一串具有竖直整数边的矩形，它们以某种方式“合成”大矩形的一条水平整数边或竖直整数边。

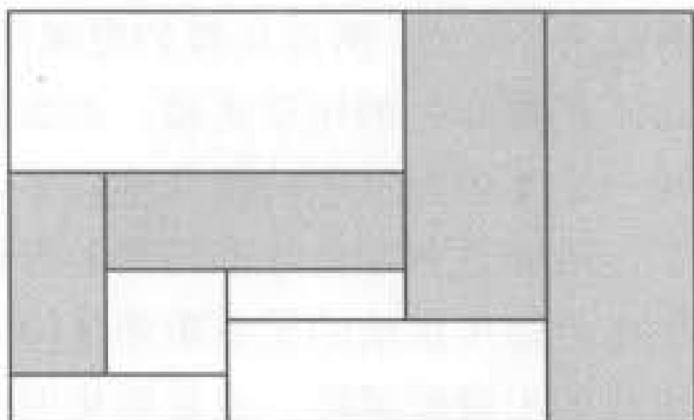


图 5

这里，我们需要对于每种可能的分割都找到这样一个矩形链。但是分割是非常难以处理的，而且每个矩形的整数边可能是水平的或竖直的，某些矩形可能同时有这两种整数边。那么我们怎样才能找到一种合适的方法来描述和处理所有这些可能性呢？

这些矩形链到底是如何起作用的呢？如果若干个小矩形都有一条水平整数边，并且如图 5 所示，它们之间是从一个个小矩形的竖直边连接到另一个小矩形的竖直边，那么大矩形就有一条水平整数边，因为大矩形的水平边的长度恰好是这些小矩形的水平边的长度之和。(换句话说，如果你把若干积木一块块摞起来，这一构造的总高度就是每块积木的高度之和。)[②]

寻找这样的矩形链的部分困难在于我们不知道哪些小矩形有水平整数边，而哪些小矩形有竖直整数边。为了包括各种可能性，想象凡是有水平整数边的小矩形都被染成绿色，

而凡是有竖直整数边的小矩形都被染成红色。(当然，既有水平整数边又有竖直整数边的小矩形更好处理：它们可以被指定染成两种颜色之一。)这样每一个小矩形就被染成绿色或红色。现在我们需要找到一条连接大矩形的两条竖直边的绿色矩形链或一条连接两条水平边的红色矩形链。

直接证明看起来不可行，所以让我们尝试用反证法证明。假设两条竖直边不能被绿色的矩形连接。那么它们为什么不能被连接呢？这一定是因为绿色的矩形不够多，而被那些红色的矩形阻断了。阻断这些绿色矩形与两竖直边相连的唯一可能是有一个由红色矩形组成的无缝隙的屏障，而这样的一个由红色矩形组成的无缝隙屏障一定是连接两条水平边的。所以，要么有绿色矩形链连接两条竖直边，要么有红色矩形链连接两条水平边。(熟悉“六角棋”(Hex)●游戏的人可以看出这里的相似之处。)

(顺便提一句，虽然上一段陈述的原理很直观，但要以严格的拓扑学方法证明它还需要做一些解释。简而言之，由全部绿色区域组成的集合可以被分成相连的子集合。假设其中不存在连接两条竖直边的子集合，考虑与左边的竖直边连接的所有绿色相连的子集合的并，那么与这个并集合的外部边界邻接的一个小条形区域将被染成红色，而这个红色条形区域就可确定一个连接两条水平边的红色矩形链。)[23]

现在还有一个小问题要说明，即验证连接两竖直边的绿色矩形链的确能够保证大矩形有一条水平整数边。这里真正的问题仅是：链中有多余的矩形(这是容易把它去掉的)；那

● 译者注：“六角棋”是一种益智休闲游戏。棋盘由 11×11 的六角形网格组成，玩法类似中国的围棋，两个对手中先得到连接两边的棋子链者为胜。在互联网上有该游戏的具体玩法，网址：http://mathworld.wolfram.com/GameofHex.html



些仅在角上邻接的矩形也不是问题；而链中的返向部分也很容易处理(我们只需减去而不是加上相应的整数边，那么边长总和将总是一个整数)[23]。

> **问题 5.3** (Taylor，1989，第 8 页)　平面上一个有限点集，其中任意三点均不共线。某些点之间由线段相连，但每个点最多在一条连线上。我们施行以下的操作：取两条相交的线段，例如 AB 和 CD，然后去掉它们并用线段 AC 和 BD 替代。试问：这种操作能否无限地进行下去？

首先，我们应确保这种操作不会导致退化或有歧义的情形，特别是我们不希望出现长度为零的线段或两条线段重合的情况。这就是为什么我们给出“每个点最多在一条连线上”的条件。总之，这很容易被证实，但还是应予以考虑。(有时，它可能变成一个很棘手的问题！)

通过尝试一些例子，这个结论似乎是合理的：在若干次操作后，所有线段逐渐地被转换为更靠外的线段且不再相交。但这只是口头上的描述，我们应如何用数学语言来表述呢？

我们需要以某种方式来刻画这个系统的“边的靠外性”在每次执行操作后都会增加。但是因为适用于这个系统的几何构形只有有限种，所以不可能无限增加，最后这种“边的靠外性”应达到最大值，操作过程也就停止了。(也就是说，事情在不能继续进行操作时就该结束了。)

所以，现在我们需要做以下工作：

(a) 找出系统的某一特征，它可以用数值来表示，例如，交点数、线段数，或者精心设计的点的分数(就像飞镖游戏中的计分方法一样)。它必须能反映出“边的靠外性”。也就是

说，当所有的线段都被分散到边缘时，这个数也应变得更大。

(b) 这一特征应在每次操作后增强(或者维持不变，但这种特征就要弱很多)。

(例如，熟悉“豆芽”(Sprouts)● 游戏的人可能会注意到，每进行一步(连接两个点并在连接线段中放置第三个点)，有效的出口数(出口是指从某个点发出的未用过的边；游戏开始时每个点都有三个出口)就减少一个(一条线段要用掉两个出口，而一个新的点又会产生一个出口)。这表明游戏不能永远持续下去，因为出口总有用完的时候。)

现在我们需要找到一个满足(a)和(b)的特征。因为有若干个特征都满足(a)和(b)，所以不存在唯一解。不过我们只需要一个。最好的方法是去估计某些简单的特征，但愿它行得通。

让我们先从最简单的试起。“点顶数”有用吗？因为它从来不变，所以用不上。“线段数”出于同样的原因也用不上。“交点数”看起来也许用得着，但“交点数”在每次操作后并不总是减小的(虽然它最终应减小)，这一点可以从图 6 看出：一个交点变成了三个。

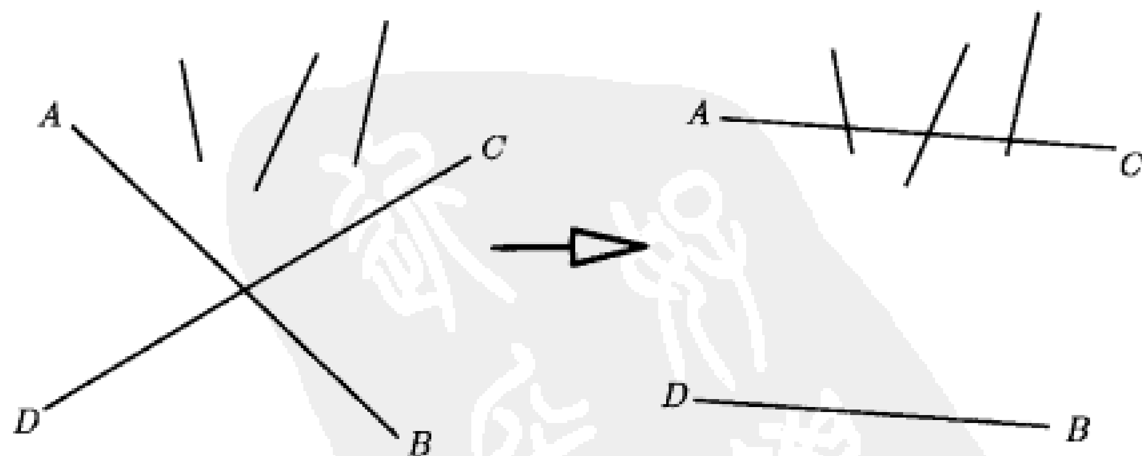


图 6

● 译者注：“豆芽”是由英国剑桥大学教授 J. H. Conway 与其学生共同发明的一种纸笔游戏。在互联网上有这一游戏的具体玩法，网址：http://en.wikipedia.org/wiki/Sprouts_(game)



当两条相交的线段变成两条没有交点的线段时，到底是什么在减少呢？由于某种原因，这些线段以某种方式变得更加分散了。从这一点看，也许可以尝试用“线段间距之和”之类的参数来描述，但这处理起来不太容易。不过沿着类似的思路，我们最后会自然地想到“线段的长度之和”，因为它不仅使线段在每次操作后变得彼此更加分离，而且也变得更短了。(三角不等式——三角形的两边之和总是大于第三边——十分漂亮地证明了这一性质。)这意味着每次操作后所有边长的总和势必减小，因此这种操作不能循环或永远进行下去(因为这些点是固定的，所以连接这些固定点的线段只有有限种可能性)，于是问题解决了。

因为在每次操作中我们改变两条线段，所以任何所考虑的特征与这些单个的线段有关，要比与交点或其他属性有关要好。这些单个的线段实际上只有三种属性：长度、位置和方向。位置和方向类的属性不能给出好的结果，因为这些属性不能真正地减少或增加。例如，要求每次操作使得总体方向(无论它的具体含义是什么)按顺时针转动似乎是不可能的。否则，为什么是顺时针而不是逆时针？顺时针与逆时针之间并没有真正的区别，但长与短之间却是明确不同的。鉴于这种考虑，我们不得不利用“总长度”这一想法。

问题 5.4 (Taylor，1989，第 34 页，问题 2) 一个男生在正方形游泳池中央，而他的老师(不会游泳)站在游泳池边的一个角上。老师的奔跑速度是男生游泳速度的 3 倍，但男生比老师跑得快。男生能逃脱老师的追逐吗？(假设两人都可以自由移动。)

让我们先画一张示意图 (图 7) 并标上已知的点：假设男

生从点 O 出发，老师从一个角出发，例如角 A。我们也可选择单位长度作为游泳池的边长。

为了求解这个问题，我们应先猜测答案会是什么。(如果你不知道自己在找什么，就不可能真正找到一种解法。)依照情形来看答案有点儿不太确定：如果男生可能逃脱，那么他一定有一个赢的计策；否则，老师就有一个赢的策略，使得不管男生如何逃跑，他总是可以通过某种方式巧妙地截住男生。后一种可能性从数学上来看有点儿苛刻：我们需要找到一种策略，能阻止男生所有可能的移动，而对于男生来说，他有许多许多种选择(他可以在二维空间中移动，而老师被严格地限制在一维空间中)。然而第一种可能性比较容易处理，我们只需设计一种巧妙的策略，然后证明它可以成功，而不用太多的反复试验。显然，证明某种策略行得通比证明所有其他策略都行不通要容易。因此，让我们假设男生可以逃脱。看起来，这是两种选择中较容易处理的情况。我们通常是先处理简单易行的选择，这样你也许能够避免一些艰难的工作。(这不是懒惰，而是一种实用性。只要能完成任务，容易的方法当然比难的方法要好。)

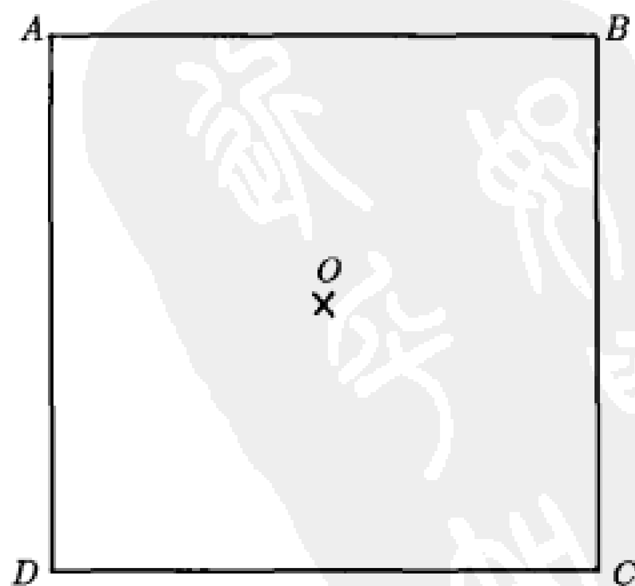


图 7



男生比老师跑得快，这意味着一旦男生上了岸并且没被老师逮住，他就能逃脱了。因此他的首要目标就是离开游泳池。在指出这一点后，男生的奔跑速度就不重要了。

在开始设计策略之前，让我们先用基本常识来排除一些不可行的策略，并把一些有可能取得成功的策略分离出来。首先，男生应以他的最快速度移动，因为即使他通过减速可能获得某些微弱的优势，老师也可以简单地通过减速与之步调一致。同样地，停下来也是无用的，因为老师也可以停下来，直到男生再次移动。(从男生的角度来讲僵局并不是一种胜利。)其次，我们可以假设老师不是容易应付的对手，他将坚守在池边。(为什么要离开池边？那只会使老师变慢！)再次，既然男生试图尽快到达池边(或者说，至少要比老师更快地达到那里)，所以沿直线移动(其距离最短，所以最快)很可能是答案的一部分，虽然辗转穿梭也可能是男生用来取得优势的办法之一。最后，策略不应完全事先确定，而应在一定程度上依赖于老师的行动：因为，如果老师知道男生将蛇行片刻直到池角，比如说，角 B，而男生仍然按照他事先确定的这种相当愚蠢的计划行动的话，那么老师就可以直接跑到角 B 等着男生的到来。

综上所述，男生的最佳策略就是以最快的速度沿直线猛冲，同时根据老师的行动灵活地改变其策略。

牢记这些一般性的准则，我们就可以开始尝试一些策略。显而易见，男生应向远离老师的方向移动，因此直奔角 A 就不是明智之举。根据直觉，应沿着直线跑向距离角 A 最远的角 C。这时男生需要游过 $\sqrt{2}/2\approx0.707$ 个单位长度，而老师需要从角 A 跑到角 B 再跑到角 $C(A\rightarrow B\rightarrow C)$或从角 A 跑到角 D 再跑到角 $C(A\rightarrow D\rightarrow C)$，即跑过游泳池的两个边长，才

能到达男生上岸处。老师的速度是男生的 3 倍，所以当男生仅游完 2/3≈0.667 个单位长度时，老师就已到达角 C。因为老师先到，所以这种方法行不通。

与其一味地跑，不如采取机动方案。毕竟，可以设法从池边的任何一处离开游泳池，而不一定是池角。例如，让我们试图朝着角 B 和 C 的中点 M 游，那么男生只需游过 1/2=0.5 个单位长度，但是老师也不必跑那么远了($A\rightarrow B\rightarrow M$ 是 1.5 个单位距离)。由于老师的速度是男生的 3 倍，所以当男生爬出游泳池时老师将恰好抓住他。

当老师沿着池边跑时，他几乎让男生逃脱，因为如果老师稍慢一点儿，那么男生就能逃脱。这表明：

- 老师的速度是恰好能够截住男生所需的最小值；
- 老师的速度是恰好能够使男生得以逃脱的最大值。

这使问题变得有点儿复杂，因为这里老师的速度似乎处于临界状态。如果老师的速度稍慢一点儿，男生只要径直朝着池边游就可以逃脱。如果老师的速度快很多，则只需跟着男生跑就行，如当男生做顺时针移动时老师也随之按顺时针方向跑，等等。常识无法给出一个定论，我们还需要做一些计算。

如果男生朝一条边游去，那么老师就需要不停地跑才能刚刚跟上他。换句话说，男生只要预示朝某个方向移动就可以强迫老师移动。由于在某种程度上能控制局势，男生的这种移动主导权就可能是一个强有力的工具。我们能否运用它呢？

假设男生正在朝着 BC 的中点 M 全速前进，老师除了先跑向角 B 然后奔向 M 点外，没有其他选择。如果老师改变方向，或者采取其他行动，男生就可继续前进，并在老师之前到达池边。但是男生不必一路游到岸边，他只要这样威胁一下就可使老师跑下去。其结果是男生可以促使局势发展到这



样一种状态(图 8)：

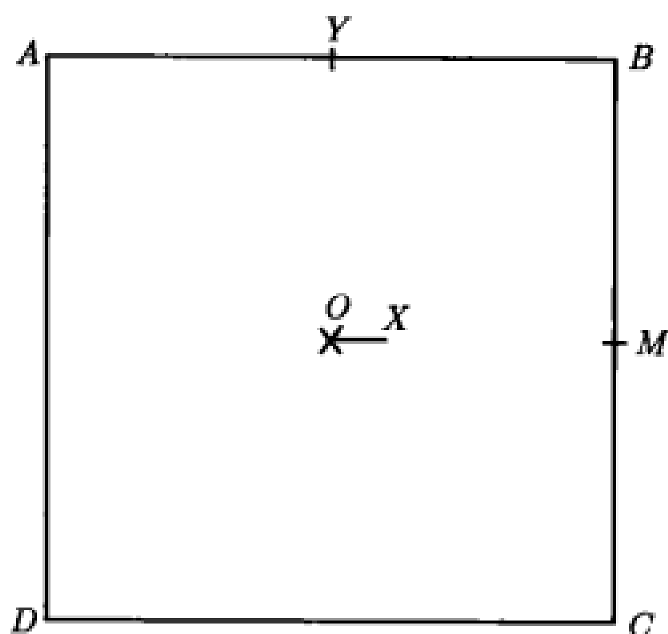


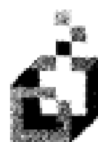
图 8

如果男生全速游到中间某点 X，老师一定在点 Y 处(Y 满足$|AY|=3|OX|$)。因为在男生到达点 X 之前，老师没有足够快的速度跑过 Y 点，而且如果老师还没到达点 Y，男生只需继续径直游向点 M 就可以摆脱老师，所以老师不得不到达点 Y。

现在的情况是：男生在点 X 处，老师被迫到达点 Y。男生还有一直向点 M 前进的必要吗？向点 M 前进的威胁足以使老师到了他现在的地方。但威胁和现实是不同的，既然老师在边 AB 上束手无策，男生为什么不向对边 CD 冲过去呢？到达岸边只需半个单位长度，而且和第一次我们考虑男生冲向池边不同的是，这时老师处于不利的位置。事实上，如果从点 O 到点 X 的距离小于 1/4 单位长度，那么很容易看出老师因距离太远而不能逮到男生。这样，男生就十分容易逃脱了。

> *习题 5.2 假设老师的奔跑速度是男生游泳速度的 6 倍。证明男生不能逃脱。(提示：画一个假想的正方形，边长为 1/6 个单位长度，中心在点 O 处。一旦男生离开这个正方形，老师就会占上风。)

**习题 5.3 假设游泳池是圆形，而不是正方形。很显然，男生只要朝着与老师相对着的点游去就可以逃脱。但是如果老师跑得更快会怎样呢？更确切地讲，老师能够逮住男生所需的最慢速度是多少？为了找到一个下界(也就是说，为男生设计一个逃跑策略)及算出一个上界(这需要为老师设计一套如何移动的完备策略)需要具有双倍的创造性(或者变分学的知识)。(对于正方形游泳池也可以问相同的问题，但比圆形的问题更需要技巧。)



第六章　其 他 例 题

数学有时被看做一个庞大的体系，就像一棵树，分成几个大分支，每个分支自身又分成各个专门的领域，到了树的末端你才可以看到花簇和果实。

但是对整个数学进行如此条理清晰的分类并不是一件容易的事。在各个分支之间总有一些模糊的领域，同时还有一些在所有经典分支以外的特殊领域。

以下几个问题十分有趣，它们既不完全属于博弈论或组合学，也不完全属于线性规划的范畴。

> **问题 6.1** (Taylor，1989，第 25 页，问题 5)　假设在某岛上生活着 13 只灰色的变色龙、15 只棕色的变色龙和 17 只深红色的变色龙。如果两只不同颜色的变色龙相遇，它们就都变成第三种颜色(例如，一只棕色的变色龙和一只深红色的变色龙相遇时两者就都变成灰色的变色龙)，且这是它们唯一的变色机会。试问：所有的变色龙能否最终都变成同一种颜色？

题目中的“最终”一词带有开放式问题的味道。这意味着我们需要判定在所有可能的变色龙颜色组合中是否包括一种所有变色龙具有单一颜色的情形。

从启发式思考的角度来看，我们应先尝试否定答案。如果答案是肯定的，那么就应该有一种特定的步骤来实现这个目标。这听起来更像一个计算问题，而没有多少数学味道，但该问题出自于数学竞赛，于是我们有理由相信答案是否定的。因此我们还是先来试图证明否定的答案。

为了证明答案是否定的，弄清楚哪些状态可以实现，而哪些大概不能实现，也许是一个好主意。一旦我们找到了某种规律，下一步的证明就有了一个明确的目标。正如我们在前几章所看到的，为了解决一个数学问题，你经常需要猜测某个中间结论，此中间结论可以推出要证的结论，但两者在逻辑上并不等价。虽然从逻辑的角度看这将带给你一个可能更难证明的问题，但是从实用的角度看它会提供一个与给定信息更为接近的目标，并使我们专注地向一个更明确的方向努力。此外，对结论的推广也可能剔除某些表面信息，取得更多的收获。

举一个简单的例子，假设在一张国际象棋棋盘的顶角上摆一个象(象沿着对角线移动)，我们要证明它绝不可能移动到相邻的顶角处(即不与这个顶角相对的两个顶角中的任一个)。代替证明这个结论，我们可以证明更一般的结论“象一定移动到相同颜色的方格内”(国际象棋棋盘是由黑白相间的方格组成的)。从逻辑上讲，还可以证明更多的结论，但是现在很容易看出如何继续推理下去：象的每一步移动都使之保持在相同颜色的方格内，因此不管移动多少步都不会离开相同颜色的方格。

无论如何，我们需要先引进某种合适的符号(例如数字和等式)。在任意给定的时刻，唯一的重要信息是多少变色龙是灰色的，多少是棕色的，以及多少是深红色的(问题的设计不允许变色龙呈现任何其他颜色)。我们可以利用一个三维向量有效地表示这一信息：变色龙的初始状态是(13,15,17)，而题目问的是我们能否通过改变颜色的操作来达到(45,0,0)，(0,45,0)或(0,0,45)的状态。改变颜色的操作是这样构成的：从其中的两个坐标上各减去 1 并给第三个坐标加上 2。这样



我们就得到了一个向量表达式，这是一个真正解决问题的方法。

(这里我们给出证明的简要轮廓：设向量 $\boldsymbol{a}=(-1,-1,2)$，$\boldsymbol{b}=(-1,2,-1)$和 $\boldsymbol{c}=(2,-1,-1)$，那么两只变色龙的相遇可表示为把向量 $\boldsymbol{a}$, $\boldsymbol{b}$ 或 $\boldsymbol{c}$ 中的一个加到当前的“状态向量”上。因此，这样的做法可以达到的任意状态必定为如下形式：$(13,15,17)+l\boldsymbol{a}+m\boldsymbol{b}+n\boldsymbol{c}$，这里 l, m 和 n 是整数。于是你所需要证的是，向量(45,0,0)不能表示成以上形式。这在克拉默(Cramer)法则中，或者初等的整数运算中，是一件很简单的事情。)

让我们尝试找一种更优美的方法，如前面所概述的那样，来找出变色龙所有可能的颜色组合。首先，变色龙的总数一定保持不变。但是这在此问题中用处不大(虽然考虑总数有时在类似的问题中可能是一个好主意)。其次，两只不同颜色的变色龙相“融合”就变成另一种颜色。我们可以重点考虑这种融合现象。比如说，当两个水面高度不同的水容器在其底部相互连通时，水位就“融合”到中间位置，但是水的总量保持不变。因此我们能否说“颜色总量”保持恒定呢？

显然，我们需要定义“颜色总量”使其适用于数学。以一只灰色的变色龙和一只深红色的变色龙“融合”成两只棕色的变色龙为例。如果我们设，比如说，灰色的色值为 0，棕色的色值为 1，而深红色的色值为 2，那么这里的“颜色总量”就保持不变(一个“0”和一个“2”合成两个“1”)。但是，比如说，当试图“融合”一只深红色的变色龙和一只棕色的变色龙时，这一点却不成立了。看起来没有一个记分体系能够满足“融合”的所有三种(或者甚至两种)可能性。

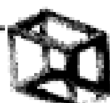
问题的症结在于这种“融合”过程的执行具有循环性。但是不要彻底放弃！取得部分成功(或部分失败)的尝试很可

能是一种真正成功方法的组成部分。(那么反过来，对于微不足道的成功也不必太激动。)考虑光学中的三原色：红色、蓝色和绿色。如果让一束红光和一束绿光同时发出，我们就得到一束具有双倍亮度的紫红色光，也就是，一束非蓝色的光。这三种原始颜色也是循环的。我们能否通过某种方式利用这种色光原理来类比我们的问题呢？

其实，两者唯一的本质不同是，对于光束，红色和绿色合成非蓝色，而不是蓝色。但是，等一等！我们可以利用模算术的方法使蓝色等效于非蓝色。基于这种想法，可以尝试对模 2 来考察我们的向量：我们的向量由(1,1,1)开始，而且一定要避免它变成(1,0,0), (0,1,0)或(0,0,1)。遗憾的是，这样行不通。但是现在妖怪已经钻出了瓶子[●][㉘]，我们可以尝试其他模数。马上想到的模数是 3(毕竟是有三种循环的颜色)。现在我们可以试图用如下两种策略中的任何一种来攻克这个问题：

- 向量方法：我们的初始向量(13,15,17)现在变为(1,0,2) (mod 3)，而且研究结果显示颜色的改变只能使得向量变成(1,0,2)，(0,1,2)和(1,2,0)[㉙]，绝不可能产生三个目标向量(45,0,0),(0,45,0)和(0,0,45)中的任何一个，因为它们都等于(0,0,0) (mod 3)。
- 颜色总量方法：我们原来计算“颜色总量”的方法是给每种色值指定一个数。既然我们知道了模数，为什么不用模算术中的色值呢？假设灰色的色值为 0(mod 3)，棕色的色值为 1(mod 3)，而深红色的色值为 2(mod 3)。这样，就使得总色值一定保持不变(因为三种可能的融合都不会引起总

● 译者注：这一典故出自《一千零一夜》中的“阿拉丁和神灯的故事”，意思是妖怪一旦跳出瓶子，就可以施展它们的魔法。



色值的变化——自己试试看吧)。初始的总色值是 $13\times0+15\times1+17\times2=1 \pmod 3$，而我们的三个目标的色值(灰色为 45，棕色为 45，或者深红色为 45)都是 $0 \pmod 3$。

习题 6.1　六位音乐家一同参加一届音乐节。在每场音乐会上，有些音乐家演奏，而其他音乐家作为观众欣赏。至少需要安排多少场音乐会，才能使得每位音乐家可以作为观众欣赏到其他所有音乐家的演出？(提示：很显然，在一场音乐会上，并非每一位音乐家都能欣赏到其他所有音乐家的演奏，所以为了实现所有“欣赏的可能性”，就需要一场以上的音乐会……按照这一思路考虑，再引进一种合适的“记分”方法，你将得到一个所需音乐会场数的合理的下界。然后找出一个满足这个下界的例子，问题就解决了。)

习题 6.2　三只蝗虫在一条直线上，每一秒钟，都有一只蝗虫(且只能是一只)跳过另一只蝗虫。证明：在 1985 秒后这三只蝗虫不可能是开始时的排列次序。

习题 6.3　假设有四枚跳棋棋子被摆放成一个边长为 1 的正方形。假设你走棋的次数不受限制，每次走棋时，举起一枚棋子跳过另一枚选定的棋子到一个新位置，使得被跳过的棋子与这枚棋子的新位置和原位置的距离相等(当然，方向相反)。此外，对于这两枚棋子距离多远才可以这样跳没有限制。试问：能否移动这些棋子使得它们被摆放成一个边长为 2 的正方形？(如果你的思考恰好对路，就会有一种非常完美的解法。)

*问题 6.2　爱丽丝、贝蒂和卡罗尔三个女孩参加同一系列考试。在每门考试中都有一个女孩的分数为 x，另一个女孩的分数为 y，第三个女孩的分数为 z，这里 x, y, z 是不同的正整数。所有的考试结束后，爱丽丝的总分是 20，贝蒂的总分是 10，卡罗尔的总分是 9。如果贝蒂的代数成绩排名第一，那么谁的几何成绩排名第二呢？

在这个问题中只有一点儿经过加工的信息，我们所知道的似乎只有总分。那么如何由总分确定单科考试的分数呢？因为我们有其他信息可以利用，所以也许可以做到这一点。在每门考试(我们并不知道到底有几门考试)中都有一个女孩得分为 x，一个女孩得分为 y 及一个女孩得分为 z。这是一条独特的信息，我们该如何利用呢？

首先，我们可以试图将它与第三条信息，即贝蒂的代数成绩排名第一配合使用。这就意味着贝蒂的分数是 x, y, z 三种选择中的最高分。为了简单起见，设 x 最大，而 z 最小，也就是说，$x>y>z$(请记住已知 x, y 和 z 是不同的)。这样我们并没有丢失什么信息，但却使问题得到简化：我们可以说贝蒂在代数考试中得了 x 分。

但是对于其他科考试，我们仍然对她们得分的各种可能性了解不多。例如，在几何考试中，可能是爱丽丝得 z 分，贝蒂得 x 分，卡罗尔得 y 分；也可能是爱丽丝得 x 分，贝蒂 y 分，卡罗尔得 z 分。在所有这些可能性中哪个量是不变的呢？啊——每门考试的总分保持不变。无论如何分配 x, y 和 z，每门考试的总分总是 $x+y+z$。关于这些总分，我们还能得到什么信息呢？对了，我们还知道所有考试的总分是



$20+10+9=39$，于是得到

$$N(x+y+z)=39,$$

这里 N 为考试的门数。之前我们对考试门数几乎一无所知，而现在有了这样一个关于考试门数的公式，这将有助于我们的推理。

但是，单独一个等式是不够的。然而，我们一定要牢记 N, x, y, z 是正整数，而不是实数。而且，我们还有第四条信息：x, y 和 z 是不同的数。这些条件将减少上式中的可能性。

既然我们知道 N, x, y 和 z 是正整数，那么上面的等式具有以下形式：

$$(\text{正整数 A})\times(\text{正整数 B})=39。$$

所以 N 和 $x+y+z$ 一定是 39 的因子。但是 39 的因子只有 1, 3, 13 和 39，因此我们有四种可能性：

(a) $N=1$ 和 $x+y+z=39$；

(b) $N=3$ 和 $x+y+z=13$；

(c) $N=13$ 和 $x+y+z=3$；

(d) $N=39$ 和 $x+y+z=1$。

但是并非所有这些可能性都是合理的。例如，可能性(a)表示只有一门考试。这与问题的语义有所冲突，题目中暗示至少有两门考试(代数和几何)。可能性(c)和(d)，除了考试门数看起来多得让人受不了外，因为 x, y, z 是不同的正整数(这要求 $x+y+z$ 至少是 6)，所以也都行不通。因此唯一没有被排除的可能性只有(b)；这样一定是进行了三门考试，且 $x+y+z=13$。

现在我们的可能性少了很多。但是我们始终不知道两条很重要的信息：x, y 和 z 的准确数值及每个女孩每门考试的得分。第一个问题可以通过 x, y 和 z 是不同的正整数且相加

等于 13 这一事实部分地得到解决；而第二个问题可以由贝蒂在代数考试中得 x 分的事实给出部分的回答。我们如何才能对这些不完整的结果加以改进呢？

每个女孩的总分这一信息还没有得到充分利用。观察题目给出的这组数据，我们看出爱丽丝的总分比贝蒂和卡罗尔的总分高很多。这就提示我们，她可能在每门考试中都得到高分(即 x 和 y)。但是贝蒂在一门考试中排名第一，所以爱丽丝的分数不可能都是 x。在最好的情形下她可能得到两个 x 分和一个 y 分。类似地，卡罗尔不太可能在任何一门考试中得到最高分 x，而且更有可能大部分分数是 z。我们能否把这些推测转化成严格的数学式子呢？

最初的回答是“也许吧”。让我们以爱丽丝的得分为例。最好的情形是她得到 $2x+y$ 分。也许我们可以证明她的得分恰好是 $2x+y$，毕竟爱丽丝的总分数比其他女孩高很多(20 比 10 或 9 大得多)。爱丽丝的分数还有什么其他可能性呢？它们是 $2x+z$, $x+2y$, $x+y+z$, $x+2z$, $3y$, $2y+z$, $y+2z$ 和 $3z$。所列的最后几个分数看来过低不可能达到 20 分。如果运气好的话，它们可能被排除掉。但是为了严格地证明这一点，我们需要 x, y 和 z 的一个适当的上界。这就是我们接下来的任务：限制 x, y 和 z 的大小，使得我们可以排除若干种可能性。

我们所知道的只有 x, y, z 是整数，$x>y>z$ 以及 $x+y+z=13$。但是这些足以使我们得到 x, y 和 z 的相当好的界限。例如，让我们来处理 z。z 不可能太大，否则 x 和 y 也将变得很大，那么就可能迫使 $x+y+z$ 比 13 大。更确切地说，y 至少等于 $z+1$，而 x 至少等于 $z+2$，故

$$13=x+y+z\geqslant(z+2)+(z+1)+z=3z+3,$$

这就要求 $z\leqslant3$。由于有一种可行组合是 $x=6$, $y=4$, $z=3$，所



以在没有进一步信息的情形下，$z\leqslant 3$ 是我们能得到的最好上界。

现在让我们来处理 y。我们可以做与上面类似的讨论，利用 $y+1$ 来约束 x。对于 z 我们只能得到下界 1。但是这就足够了：我们可得到

$$13=x+y+z\geqslant(y+1)+y+1=2y+2,$$

所以 $y\leqslant 5$。考虑到有可行的组合 $x=7, y=5, z=1$，所以 $y\leqslant 5$ 是能得到的最好上界。最后，我们可以给出 x 的上界：z 至少是 1，y 至少是 2，故 $13=x+y+z\geqslant x+2+1$，所以 $x\leqslant 10$。而且由于我们有 $x=10, y=2, z=1$，这也是可能的最好上界。

这样我们就得到 $z\leqslant 3$, $y\leqslant 5$ 和 $x\leqslant 10$。但是我们还可以做得更好。请记住贝蒂得到一个 x 分和两个其他分数。由于贝蒂的总分只有 10，所以 x 不可能等于 10。因为 $x=10$ 将意味着贝蒂在其他两门考试中得到零分，这是不可能的(我们知道所有的分数都是正整数)。事实上，x 也不可能等于 9，否则贝蒂在其他两门考试中总共只得到 1 分，这就意味着有一门考试得 0 分，这又是一个矛盾。所以我们得到 $x\leqslant 8$ 的事实。现在我们可以排除更多可能性。实际上，容易看出爱丽丝唯一可能的总分是 $2x+y$，所有其他的分数都不可能达到 20。例如，$2x+z$ 最多为 $2\times 8+3=19$。

因此爱丽丝得到两个 x 分和一个 y 分。既然贝蒂在代数考试中得了 x 分，爱丽丝在这门考试中一定得了 y 分。我们可以把这一信息，连同所知道的其他信息整理成一张表(表 1)。

表　1

	爱丽丝	贝蒂	卡罗尔	总和
代数	y	x	?	13
几何	x	?	?	13
其他	x	?	?	13
总和	20	10	9	39

我们现在就可以看出卡罗尔在代数考试中一定得了 z 分，因为这是唯一剩下的分数了。

我们离目标越来越近了，已经知道在几何考试中以 y 分排名第二的是贝蒂或卡罗尔。但是我们还确定不了到底是谁。观察表 1 中爱丽丝的一列，我们又得到了一条信息，即 $y+x+x=20$。记起 $x>y$ 和 $x\leqslant 8$，所以只有两组解：$x=8$, $y=4$ 或 $x=7$, $y=6$。但是因为 $x+y+z=13$，所以不可能有 $x=7$ 和 $y=6$，否则将造成 $z=0$。所以只可能有 $x=8$, $y=4$，这就使得 $z=1$。这样我们就完全确定了 x, y, z 的取值，从而取得了重大突破。我们现在可以把表 1 更新为表 2。很容易看出在几何和其他一门考试中贝蒂都需要得 z=1 分，而卡罗尔在这两门考试中都需要得 y=4 分。因此本题的答案是卡罗尔在几何考试中排名第二。

表 2

	爱丽丝	贝蒂	卡罗尔	总和
代数	4	8	1	13
几何	8	?	?	13
其他	8	?	?	13
总和	20	10	9	39

问题 6.3 (Taylor，1989，第 16 页，问题 3) 两个玩家用由 60 小块组成的一块巧克力做游戏，这块巧克力是一个 6×10 的矩形。首先第一个玩家沿着划分巧克力块的浅槽掰下一部分，并把掰下的部分扔掉(或吃掉)。然后第二个玩家在剩下的巧克力块上再这样掰下一部分并扔掉。游戏不断继续下去，直到只剩下一小块巧克力为止。谁能给对手留下单独一小块巧克力，谁就是赢家(即最后一个掰巧克力的人)。试问：哪个玩家有完美的获胜策略呢？



顺便提一句，容易证明，对于任何步数有限的智力游戏，其中某位玩家一定有某种取得胜利(或平局)的策略。这可以通过对游戏的最大步数做归纳法来证明。即使国际象棋也受到这样的限制，虽然现在还没有人找到这个策略，大多数人都认为它实在是太复杂了。既然在这个游戏中不存在平局，其中一位玩家必定有一种保证获胜的策略。但是谁是赢家呢？

首先，让我们把问题从巧克力转化到数学上。我们可以从规范掰巧克力块的过程入手。掰过巧克力的人都知道，掰巧克力的唯一方式是把它掰成两个矩形，而不是锯齿形的或不完整的矩形。实际上，我们把一块 6×10 的巧克力掰成一个小的矩形，那么小矩形和原来的矩形一定有一个长或者宽是相等的(观察图 1，其中虚线是掰开的地方)，也就是说，巧克力被掰成宽度相等但长度较短或长度相等但宽度较窄的一块。例如，在图 1 中，6×10 的巧克力被掰成 6×7 的一块(6×3 的小块被扔掉或吃掉了)。

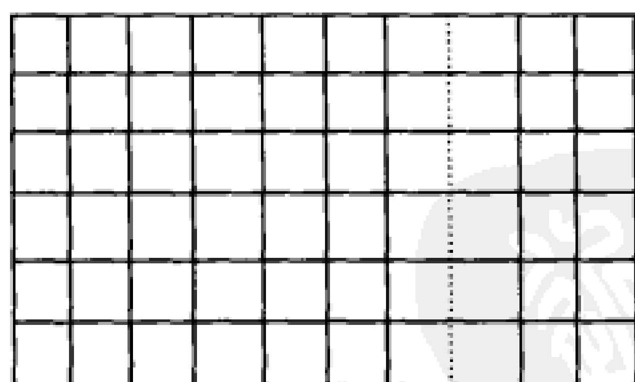


图　1

现在我们需要对这个矩形引进一些符号，最好是数字符号。我们该如何用数字来表示这块矩形巧克力呢？显而易见的方案是给出块的长和宽。所以初始时的巧克力块是一个 6×10 的块，或用点坐标表示为(6,10)。巧克力块的位置无关紧要，尺寸才是重要的。我们的目标是把(1,1)留给对方，那

么我们的规则是什么呢？我们可以在横向或纵向掰下一块，当然不能是零或负的。例如，我们可以从(6,10)处移动到以下位置上：

$$(6,1),\ (6,2),\ (6,3),\ \cdots,\ (6,9),$$
$$(1,10),\ (2,10),\ \cdots,\ (4,10),\ (5,10)。$$

总之，我们可以做水平向左或竖直向下的移动。示意图 2 抽象地演示了这种移动，给出了从(6,10)处开始的两种可能达到的状态。

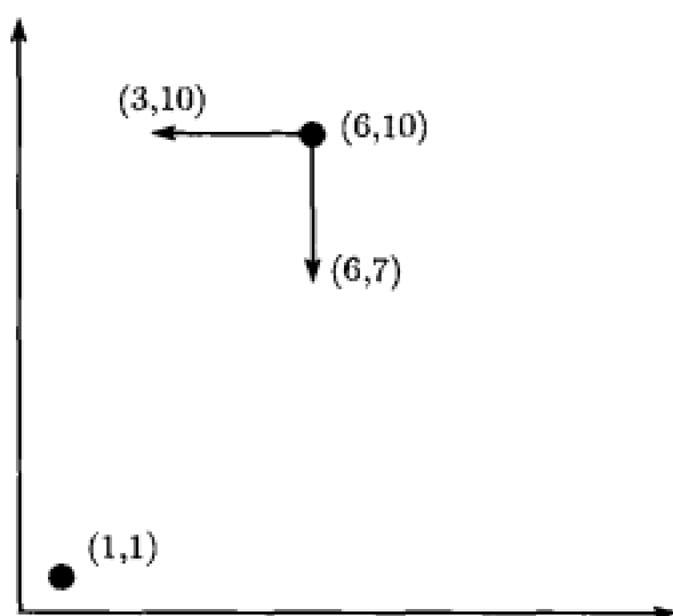
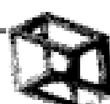


图 2

现在我们对这个巧克力问题有了一个不错的数学模型，就可以用数学的语言(不过没那么美味可口了)重新叙述这个问题如下：

> 两个玩家轮流在格子上移动一个点，每次移动可把这个点向左或向下移动整数格，点不能越过两条轴线中的任意一条以及点的初始位置在(6,10)处。谁把点移动到(1,1)处谁就是赢家。试问：谁有一种完美的获胜策略呢？



或者我们有另一种表述：

> 两个玩家从两排筹码上轮流取筹码。每人必须从上排或下排取，但不能同时从两排上取。开始时上排有 5 个筹码，而下排有 9 个筹码(这表示点(6,10))。取到最后一个筹码的就是赢家。谁有一种完美的获胜策略呢？

在这种表述中所做的一点儿修改是从上排和下排都减去了 1。这对于熟悉 Nim[●] 游戏的人应该是很强的暗示，他们将会很容易地解决问题。但是即使没有 Nim 游戏和与搏弈论方面的知识，我们也可以解决这个问题。

现在我们有了符号和一个抽象的数学模型，接下来需要充分理解这个问题。问题在于 6×10 的块有许多种掰法。试验应从一个较小的块开始，于是让我们从一个 2×3 的块入手。

第一个玩家可能留下 1×3, 2×2, 2×1 的块中的一个。留下 1×3 和 2×1 的块是很愚蠢的，因为第二个玩家可以就此掰去除最后一个 1×1 的块之外的所有部分从而获胜。所以第一个玩家应留下一个 2×2 的块。那么第二个玩家只能留下一个 1×2 或 2×1 的块，而第一个玩家只要掰去剩余块的一半，剩下 1×1 的块，因此赢得游戏。所以对于一个 2×3 的块，第一个玩家获胜。

我们从中没有得到很多信息，于是继续尝试另一个例子，例如一个 3×3 的块。现在，对于留下什么样的块，第一个玩家有几种选择：1×3, 2×3, 3×2 和 3×1。利用对称性可以有效地排除最后两种选择。留下 1×3 的块是很愚蠢的，因为第二个玩家可以掰去除最后一个 1×1 的块之外的所有部分

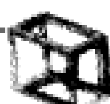
● 译者注：一种游戏。把火柴棍等物品排成行，参加者轮流拿，拿到最后一个的人赢。这一游戏是博弈论中一种经典的模型。在互联网上有这一游戏的具体玩法，网址：http://en.wikipedia.org/wiki/Nim

从而获胜。但是留下 2×3 的块同样不好，因为我们已经在上段中解决了这问题，第二个玩家可采用在上段中第一个玩家曾用过的策略：得到一个 2×2 的块，使第一个玩家除了留下一个 1×2 的块外别无选择，于是第二个玩家可以掰完后留下一个 1×1 的块从而获胜。因此对于一个 3×3 的块，第一个玩家失利。

我们通过利用 2×3 的块的问题解决了 3×3 的块的问题。这启发我们尝试用归纳法来解决一般的情况。例如，假设我们想要解决 3×4 的块的问题，而且我们已经知道在 3×1, 3×2, 1×4 和 2×4 的块的问题中赢家都是第一个玩家，而在 3×3 的块的问题中他却是一个输家。因此，第一个玩家在 3×4 的块的问题中，其策略是把一个 3×3 的块留给第二个玩家，因为对于第二个玩家这是必输的块。所以，第一个玩家的策略就是留给第二个玩家一个不得不掰、一掰就输的块。为什么这些块是必输的呢？因为无论你怎么掰，它们都变为对手必赢的块。而这些块是必赢的块，因为你可以将它掰完后留给对手一个必输的块，等等。因此我们现在的策略就是确定所有必赢块和必输块。

1×1 的块明显是一个必输块，它不能被掰，所以游戏结束。$1\times n$ 的块($n>1$)是一个必赢块，因为掰它的人可以留给对手必输的 1×1 的块。2×2 的块也是一个必输块，因为掰它的人必然会留下一个保证对手获胜的 1×2 的块。现在我们可以说 $2\times n$ 的块($n>2$)是必赢块，因为我们可以把必输的 2×2 的块留给对手，依此类推。我们注意到：

- 如果 $a\times b$ 的块是必输块，那么 $a\times c(c>b)$ 的块就是必赢块，因为掰 $a\times c$ 的块的人必然可以给对手留下 $a\times b$ 的块。例如，因为我们已经证明 3×3 的块是必输块，那么 3×4, 3×5, 3×6 的块，等等，就都是必赢块。
- $a\times b$ 的块是必输块，只有当对它的所有可能掰法留下



的都是对手的必赢块。例如，正如前面所证明的，1×4, 2×4, 3×4 的块以及由对称性而得到的 4×3, 4×2, 4×1 的块都是必赢块，因此 4×4 的块是必输块。

我们可以把这种有条有理的方法不断地进行下去，最终达到 6×10 的块。但是我们为什么不把这个过程更加数学化呢？这里应该存在必赢块和必输块的某种规律。那么目前为止哪些是我们知道的必赢块和必输块呢？我们得到的必赢块为

	1×2	1×3	1×4	1×5	…
2×1		2×3	2×4	2×5	…
3×1	3×2		3×4	3×5	…
4×1	4×2	4×3		4×5	…

而我们已经确认的必输块是 1×1, 2×2, 3×3 和 4×4。

根据以上讨论可以推测：只有 $n \times n$ 的块，即方块是必输块，而其他所有的块都是必赢块。一旦有了这种推测，我们甚至不必去证明(虽然你可以运用归纳法证明)，而只要应用它就行了。请记住我们想要的是把必输块留给对手。一旦我们猜到哪些块是必输的，就可以运用策略总是把它们强留给对手。如果策略始终有效，那当然很好；否则就是猜错了。概括地讲，如果我们的猜测是正确的，最佳谋略就是留给对手一个方块。因此这意味着对于 6×10 的块，第一个玩家有以下策略：

> 掰下部分巧克力，留下一个 6×6 的方块(第二个玩家必输)。那么无论第二个玩家怎么掰，下一步还是把它变成一个方块。例如，如果第二个玩家留下一个 6×4 的块，你就把它掰成 4×4 的方块。重复这个过程，总是给对手留下一个方块，直到最后你留给对手 1×1 的方块(使他或她认输)。

这种策略实际上很有效，因为无论对手怎么掰方块，他得到的都不是方块，而非方块很容易就可以又被变回到方块。而且因为巧克力的尺寸不断减小，所以变来变去的方块最终必然变成一个 1×1 的方块。这样，用一点儿不太严格的数学知识，我们最后得到了一种成功的策略。这正是我们要找的。

总之，这是一种解决智力游戏的标准方法：确定所有赢的状态和输的状态，然后总是向赢的状态转化。优秀的智力游戏玩家都使用这种方法，虽然他们对赢或输的状态并没有确切的判断，只有“有利的”或“不利的”推测。例如在国际象棋中，当我们说某一步是“好棋”或“臭棋”，不就是指这一步使游戏向有利或不利的局面转化吗？很少有国际象棋选手靠随意走棋而不靠试图改进局面能取胜的。

习题 6.4 两个玩家用 153 个筹码开始做游戏。每个人必须轮流每次从中取走 1～9 个筹码，取到最后一个筹码的就是赢家。第一个玩家还是第二个玩家有必胜的策略呢？如果存在这样的策略，它是什么呢？

习题 6.5 两个玩家从 n 个筹码开始做游戏。每个人必须轮流每次从中取走 d 的幂次个筹码，取到最后一个筹码的就是赢家。对于以下 d 值，确定 n 取何值时第一个玩家有一种必赢的策略；n 取何值时第二个玩家获胜：

(a) $d=2$;

(b) $d=3$;

(c) $d=4$;

*(d) 一般情形。



习题 6.6　重复以上两个练习，但是现在目标改为争取输，也就是说，迫使另一个玩家取到最后一个筹码。(如果思路正确，很容易找到答案。)

习题 6.7　考虑问题 6.3 的一种三维形式，从一个 $3 \times 6 \times 10$ 的块开始，并且沿着三维中的任意坐标轴掰。哪个玩家获胜？获胜策略是什么？

****习题 6.8**　在五子棋(Gomoku)游戏中，两个玩家(白方和黑方)轮流在一张 19×19 的棋盘上摆放一颗相应颜色的石子。如果一个玩家使自己的五颗石子摆成一行(在任何方向上)，他就获胜了；如果棋盘上的所有方格都被填满却还没有出现"五子连行"，游戏就成为平局。证明：第一个玩家有一种保证至少是平局的策略。(提示：你需要用反证法来论证。先证明如果第一个玩家不能至少保持平局，那么第二个玩家就有一种必胜的策略。然后让第一个玩家"偷用"这一策略。)

问题 6.4 (Shkarsky 等，1962，第 9 页)　两兄弟卖了一群羊，每只羊卖价的卢布数与这群羊的个数相同。然后按以下方式分配收入：哥哥先取走 10 卢布，然后弟弟取走 10 卢布，哥哥再取走 10 卢布，依此类推。分配到最后，轮到弟弟取钱，他发现已不足 10 卢布了，于是他只好把剩余的取走。为了分配公平起见，哥哥把自己的小刀给了弟弟，小刀的价格是以卢布为单位的一个整数。试问：小刀值多少钱？

我们的第一个反应可能是这个问题看起来没有给出足够的信息。其次，这个问题似乎也不够严格。但是，在进行各种尝试之前就放弃解决问题的努力，这也是不应该的。看一下问题 6.2，它可着手讨论的信息甚至更少，但仍然可以得到解决。

我们应先设法把这个问题表述为方程，为此需要引进一些变量。首先，我们注意到小刀的价格最终取决于羊的数量，而羊的数量是这里唯一一个独立变量。(也就是说，羊的数量决定一切。)让我们假设有 s 只羊，那么每只羊卖得 s 卢布，所以总收入是 s^2 卢布。

现在我们需要了解收入是如何分配的。假设总卢布数是 64，那么哥哥取走 10 卢布，然后弟弟取走 10 卢布，依此类推。这时，我们发现最后 4 卢布被哥哥取走，而不是弟弟，所以这种情况不会出现。请记住在给定信息中是弟弟最后取走现金的事实。我们如何用数学语言表达这一事实呢？

为了用数学语言来表达这一事实，我们需要一些方程和变量(以保证足以描述这一情形，但也不至于引起迷惑或造成冗余)。假设弟弟在取零钱之前已经取了 n 次 10 卢布，那么哥哥也取过 n 次 10 卢布，他还要再加上在弟弟取零钱之前又一次取走的 10 卢布，剩下的零钱只有 a 卢布 (a 是 1～9 中的某个整数，因为问题的语义似乎表明 a 是非零的)，于是总卢布数只能是

$$s^2 = 10n + 10 + 10n + a$$

或

$$s^2 = 10(2n+1) + a\text{。} \tag{19}$$

但是它与小刀有什么关系呢？我们想要求解的因变量是小刀的价格 p。我们需要一个方程联系 p 和其他变量，最好是与



独立变量 s 联系。这样在给小刀以前，哥哥取走 $10n+10$ 卢布，而弟弟取走 $10n+a$ 卢布。一旦给了小刀，哥哥的收入为 $10n+10-p$，而弟弟的收入为 $10n+a+p$。出于公平的考虑，这两部分收入一定是相等的。因此得到一个联系 p 和 a 的有用的方程：

$$a = 10 - 2p\,。 \tag{20}$$

把式(20)代入方程(19)(消去 a)，可以得到一个联系 p 与其他变量的方程：

$$s^2 = 20(n+1) - 2p\,。 \tag{21}$$

我们要利用这些方程来解出 p。因为我们不知道 s, n 或 a 的值，所以看起来信息似乎还不够充分。怎样才能进一步缩小范围呢？根本问题是这里出现的未知量太多了。我们可以利用模算术来消去其中若干未知量。例如，我们可以对式(21)取模 20 来消去 n，得到

$$s^2 = -2p \pmod{20}\,。$$

这使离求出 p 的目标更近了，但我们还要去对付烦人的 s。幸运的是，我们可以利用这样一个事实：平方数在模算术中只能取若干个限定的数值。事实上，对模 20，平方数必须取 0, 1, 4, 5, 9 或 16 中的某个值。换句话说，我们有

$$2p = 0,\ 1,\ 4,\ 5,\ 9,\ 16 \pmod{20}。$$

请注意 $2p$ 必定是偶数，因此我们有

$$p = 0,\ 2,\ 8 \pmod{10}。$$

这样我们就有了一个关于 p 的模 10 的等式，但还不能准确地确定答案。所有这些说明小刀的价格可能是 0 卢布、2 卢布、8 卢布、10 卢布、12 卢布，等等。但是小刀不可能太贵，不

是吗？毕竟，弟弟没有得到的只是 10 卢布或更少……沿着这些线索继续思考，最终你会想到 p 不仅与 n, s 有关，还与 a 有关，而 a 被限制在 1～9 之间。注意到式(1)，这意味着应该有 $0<p<5$，把这与另一关于 p 的等式相结合，就确定出小刀的价格是 2 卢布。(请注意，即使我们允许 $a=0$，这个结论仍然是对的。)

奇怪的是，尽管有足够的信息用于确定小刀的价格，却没有足够的信息来确定羊的价格或数量。事实上，关于 s 我们只能说 $s\equiv\pm4 \pmod{20}$，因此羊的数量可能是 4, 16, 24, 36, 44, 56 等。

对于这类智力难题，你需要使用所有可以获取的信息。最好的方法是把这智力难题中已知的所有信息罗列出来，并分别写出每一条信息，例如：

(a) 要分配平方数个卢布；

(b) 弟弟没有得到他应有收入的一小部分；

(c) 弟弟的这一差额由小刀来补偿。

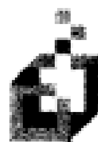
然后我们应尽快把这些事实转化为方程：

(a) $s^2=10(2n+1)+a$;

(b) $0<a<10$;

(c) $a=10-2p$。

我们应设法抓住每一条信息，无论它看起来多么无关紧要。例如，我们可以指出 n 应该是非负的，p 大概是正的(为什么问题中提到一把毫无价值的小刀)，以及有正整数只羊，等等。一旦所有信息以方程的形式确定下来，对问题进行正确处理就变得容易多了。



參考文献

书籍，就像是朋友，不一定很多，但要精挑细选。

——Samuel Paterson, Joineriana

[1] AMOC (Australian Mathematical Olympiad Committee), 1986-1987. Correspondence programme, Set 1 questions.

[2] Australian Mathematics Competition, 1984. Mathematical Olympiads: The 1984 Australian scene. Belconnen, ACT: Canberra College of Advanced Education.

[3] Australian Mathematics Competition, 1987. Mathematical Olympiads: The 1987 Australian scene. Belconnen, ACT: Canberra College of Advanced Education.

[4] Borchardt W G, 1961. A sound course in mechanics. London: Rivingston.

[5] Greitzer S L, 1978. International Mathematical Olympiads 1959-1977 (New Mathematical Library 27). Washington: Mathematical Association of America.

[6] Hajós G, Neukomm G, Surányi J, 1963. Hungarian Problem Book I, based on the Eötvös Competitions 1894-1905 (New Mathematical Library 11). Kürschák J orig comp, Rapaport E tr. Washington: Mathematical Association of America.

[7] Hardy G A, 1975. A course of pure mathematics. 10th ed. Cambridge: Cambridge University Press.

[8] Polya G, 1957. How to solve it: a new aspect of mathematical method. 2nd ed. Princeton: Princeton University. [美]波利亚 G, 2007. 怎样解题：数学思维的新方法. 涂泓，冯承天，译. 上海：上海科技教育出版社.

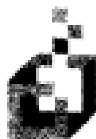
[9] Shklarsky D O, Chentzov N N, Yaglom I M, 1962. The USSR Olympiad problem book: selected problems and theorems of elementary mathematics. Suaamar I, red and ed, Maykovich J, tr.

San Franciscon: Freeman.

[10] Taylor P J, 1989. International Mathematics: tournament of the towns, questions, and solutions, tournaments 6-10 (1984-1988). Belconnen, ACT: Australian Mathematics Foundation.

[11] Thomas G B, Finney R L, 1988. Calculus and analytic geometry. Reading: Addison-Wesldy.





译　后　记

每四年举行一次的世界数学家大会于2006年8月在西班牙首都马德里举行。我有幸参加了这次盛会。这是一次万众期待的会议，因为在这个会议上将颁发具有“数学诺贝尔奖”之称的“费尔兹奖”。

这次会议的一个亮点是，被誉为“数学界的莫扎特”的澳籍华人数学家陶哲轩毫无悬念地被授予费尔兹奖。和大部分与会者一样，我也是第一次认识他，也同样地被他谦虚的态度和渊博的知识所折服。他为人平和低调，愿意和任何一位与他接触的人交流，完全没有青年天才的轻狂。在他获得费尔兹奖后，中国媒体对他进行了大量的报道，也被中国的青少年所熟悉，这里我就不赘述了。

牛津大学出版社非常适时地推出了陶哲轩在 15 岁时所写的《解题·成长·快乐——陶哲轩教你学数学》一书的修订版，并在大会期间展出和销售。当我在出版商的展台看到这本书时，立即买了下来，带回酒店阅读。我很快就被这本书所吸引，因为它既反映了一个少年数学爱好者观察问题的角度，同时也反映了作者敏锐的洞察力。书中解答数学问题的过程也十分有特色，从毫无头绪的问题开始，分析各种可能解法的利弊，排除不合适的方法，层层拨开，使题目的本质和解题技巧浮出水面，解题思路变得清晰起来。

我期望中国青少年读者可以分享阅读此书的喜悦，并从中学到若干解题的技巧。

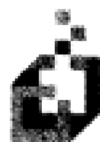
在本书翻译的过程中，南开大学组合数学中心的左连翠

老师，刘娟、白冰同学等给予了诸多帮助。北京大学出版社的孙琰和曾琬婷编辑提供了很多有益的建议，在此一并表达我真诚的感谢。潘承彪教授为本书做了校对，增强了本书的可读性和准确性，我在此除了表达我的敬意外，也一并送上我的衷心感谢。

于青林

2009年5月于济南





校后记及校注

北京大学出版社邀请我审阅本书的译文，我很高兴地接受了这一任务，因为原书是我从未见过的谈中学生如何学数学的一本好书：第一，这是被公认为“天才”、“神童”的澳籍华人数学家陶哲轩在15岁时根据自己的心得，用自己的语言所写成的；第二，在出版16年后(2006年)，作者获得了费尔兹奖；第三，在原书初版15年后(2005年底)，作者同意牛津大学出版社修订再版，并在第二版序言中着重指出他不想改变第一版的原貌，他认为这样对中学生更有益。所以，我相信，读过中译本后，无论是老师、学生，还是家长，都会从中受益，并得出自己的看法。

为了使译文尽量准确、通顺，既符合原书的风格，又适合我国中学生阅读，我仔细做了审校，提出一些修改意见，并指出了少量原书可能的疏忽(当然，也可能我领会有误，请读者指正)，供译者参考。有两处译文我请教了香港大学肖文强教授夫妇及台北九章出版社的孙文先先生，感谢他们提出了宝贵的意见。此外，为了便于读者进一步理解书中内容，我对译文加了若干校注(⓪, ①, ②, …, 见后)，其中多次引用了潘承洞和我所著的《初等数论》(第二版)(北京：北京大学出版社，2003。下同)。

潘承彪

2009年6月

⓪ 在香港大学肖文强教授的《心中有数》(台北：九章出版社，2009)一书的第198—199页，也提到了这句话，这是普罗克洛斯在评注 Euclid

的《原本》卷一时说的。但他根据的是 G. R. Morrow 的英译本(A Commentary on the First Book of Euclid's Elements, Princeton University Press, 1970)，和本书所引的英译本不同。

① 见《莎士比亚全集》(北京：人民文学出版社，1978)(第一卷)第 267 页。单数，即奇数。

② 拉格朗日定理及其证明可在很多初等数论书中找到(例如《初等数论》第六章 § 1 中的定理 1)。但是，我还没有见到过利用高斯整数(即形如 $a+bi$ 的数，其中 a,b 均为通常的整数)的有关证明。和高斯整数有关的是正整数表示为两个完全平方数之和的问题(有关知识可参看《代数数论》(第二版)(潘承洞、潘承彪著，济南：山东大学出版社，2001)中的第三章 § 3.3 以及《初等数论》第六章中的 § 2 和 § 3)。

③ 这里是指用 10 进位记数法时的末位数。书中所用的记数法都是 10 进位的。

④ 这些命题中，有的要添加适当条件后才成立，请读者自己考虑。“模算术”可参看《初等数论》第三章。

⑤ 可参看《数论导引》(华罗庚著，北京：科学出版社，1995)。

⑥ 这里的“上个世纪”是指 19 世纪。以$\pi(x)$表示不超过实数 x 的素数个数。在 1800 年左右，勒让德(A. M. Legendre)和高斯分别提出猜想：当 $x\to+\infty$时，$\pi(x)/(x/\ln x)\to 1$。1896 年，被阿达马(J. Hadamard)和瓦莱-普桑(C. J. de la Vallée-Poussin)分别独立证明。这一结论称为素数定理。关于素数定理的初等证明及有关介绍可参看《素数定理的初等证明》(潘承洞、潘承彪著，上海：上海科学技术出版社，1988)。

⑦ 本书中有关同余的知识均可参看《初等数论》第三章。几乎在所有数论书中，表示“x 和 y 相差一个 n 的倍数”的符号都用“$x\equiv y \pmod n$”，读做“x 同余于 y 模 n”，而不是这里的“$x=y \pmod n$”，读做“x 等于 y 模 n”。这样的表示式称为同余式，n 称为(同余式的)模或模数，是正整数。为了与原著保持一致，译文未做改动。在同余



式中出现的数都是整数，当 a 与模 n 互素时，同余式中的分数 $1/a$ 是表示整数 b，它满足 $ab\equiv1(\mathrm{mod}\ n)$。

⑧ 英文“digit”是指阿拉伯数字，即 0，1，2，3，4，5，6，7，8 和 9。本节讨论的是有关数的 10 进位表示中的数字问题，所以，本节的标题“Digits”可译为 10 进位表示中的数字或简译为位数。同样地，本节中把“digit expansion”译为 10 进位表示(式)；把“number of digits”译为 10 进位表示式中的数字的个数，简称位数个数，位数个数为 m 的数就称为 m 位数；把“digit-sum”译为 10 进位表示式中的各个数字之和，简称位数和；“set of digits”是指各位数字所组成的集合(出现相同的数字要重复计入)，译做位数集合；把“digits-switching”译做位数置换(或位数重排)，即把一个 10 进位数中的各位数字重新排列构成另一个 10 进位数。例如，10 进位表示式 15180 的位数个数是 5，所以它是一个 5 位数，位数和是 $1+5+1+8+0=15$，位数集合是$\{0,1,1,5,8\}$，51801 是它经过位数置换(或位数重排)得到的另一个数。

⑨ 我认为，这句话中的两个“肯定”(原文为“true”)都应改为“否定”，但未做改动。

⑩ 丢番图方程并不一定是代数方程，例如问题 2.4。

⑪ 这里用到了模 m 的剩余类、模 m 的剩余系的概念，见《初等数论》第三章§2。

⑫ 证明本题不需要用 Bertrand 假设(见《初等数论》第八章§2 中的定理 4)。本题解法见《初等数论》第一章中习题三(I)第 21 题的提示。

⑬ 符号 $\mathbf{Z}/m\mathbf{Z}$ 的定义可见《初等数论》第三章§2 中的定理 2。当 p 为素数时，$\mathbf{Z}/p\mathbf{Z}$ 的生成元是指 $\mathbf{Z}/p\mathbf{Z}$ 中的这样一个元素 g，它使得 g，g^2，g^3，…，g^{p-1} 恰好是 $\mathbf{Z}/p\mathbf{Z}$ 中除去 0 mod p 外的所有其他 $p-1$ 个元素。这样的元素 g 是一定存在的，它称为是模 p 的原根。关

于原根可参看《初等出论》第五章§1及§2中的定理2。

⑭ 强归纳法也称为第二种数学归纳法，可参看《初等出论》第一章§1中的定理4。

⑮ 为了叙述得更清楚，我们给这个多项式编号以(14′)。

⑯ 原著该题中是“$f(x)=(x-a_0)^2(x-a_1)^2\cdots(x-a_n)^2+1$”，但证明过程中表明应是“$f(x)=(x-a_1)^2\cdots(x-a_n)^2+1$”，因此我们做了更正。此外，原题中没有条件“其中 a_1，…，a_n 是两两不同的整数”，这是我们加上的。不难看出，没有这一条件，结论不一定成立。

⑰“两两不同的”是我们加上的。不难看出，没有这一条件，结论不一定成立。

⑱ 见《一个数学家的辩白》(李文林编译，南京：江苏教育出版社，1996)第24页。

⑲ 可参看《几何原本》(兰纪正、朱恩宽译，西安：陕西科学技术出版社，2003)。

⑳ 先画出 R_1 的水平中心线，R_0 的竖直中心线，R_3 的水平中心线，再画出 R_2 竖直中心线，R_0 的水平中心线，R_4 的竖直中心线，就构成这样一个图形。

㉑ 原书没有明确指出 R_2，R_3，R_4 的两边之和也都大于1，而这正是证明的关键，所以我们加以补充。这里说的“宽”和“窄”，请读者根据图形来适当理解。

㉒ 这是指关于转动惯量的平行轴定理，即质量为 m 的刚体绕某给定轴的转动惯量等于该刚体绕平行于给定轴且通过该刚体的重心的轴的转动惯量加质量 m 与两轴之间距离的平方的乘积。请读者自己据此给出物理解释。该定理可参看《新概念物理学：力学》(第二版)(赵凯华、罗蔚茵，北京：高等教育出版社，2004)第四章§4.3。

㉓ 这里的“‘合成’为……”，原文为“‘add up’ to……”。应该指出的是，用这种一串小矩形的竖直边(或水平边)“合成”大矩形的



竖直边(或水平边)不一定像由本章图 5 所给出的例子那样简单。严格地说，一串小矩形的竖直边(或水平边)“合成”大矩形的竖直边(或水平边)是指(见图 a，图 b)：在这串小矩形中，任意两个相邻矩形一定各有一条水平边(或竖直边)位于同一水平线(或竖直线)上，这样的两个小矩形称为是“竖直(或水平)相连的”；而事实上，这两个小矩形不一定如图 5 那样有公共点。当然，一般的情形可比图示更加复杂，即大矩形的竖直边(或水平边)不一定等于这一串小矩形的竖直边(或水平边)之“和”，而是它们的“代数和”。作者在本题解答的最后也指出了这一点。这些请读者自己仔细考虑。

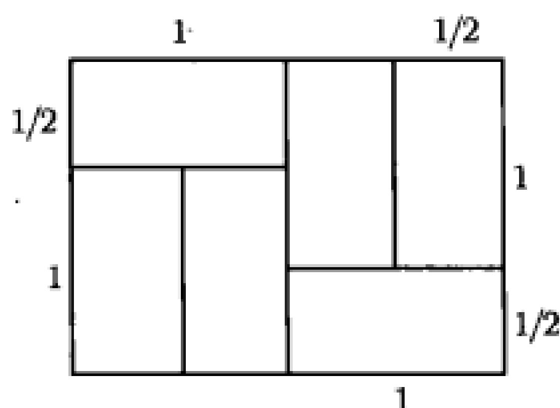


图 a

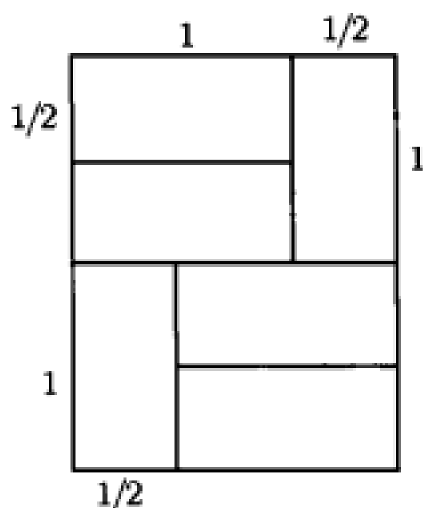


图 b

㉔《一千零一夜(第一版)》中译本(纳训译，北京：人民文学出版社，1982)，全书有六册，这典故出自该书的最后一个故事“阿拉丁和神灯的故事”(见第六册第 445—548 页)。

㉕ 这里还应该有其他几种可能，请读者自己考虑。

㉖ 这就是我国古代的筹码游戏，可参看《初等数论》附录三 § 4。